The Journal *of*
Parapsychology

Volume 11 MARCH, 1947 Number 1

EDITORIAL

PARAPSYCHOLOGY AND ETHICS

It is to ethics, the study of moral relations, that parapsychology offers its primary service. Already the psi researches have gone far enough to show the possibility of a sound and reassuring scientific approach to ethical problems which in the past have been left to dogma, speculation, and tradition. We need hardly stress the incomparably greater conviction and force which science carries today compared to these other ways of solving problems.

Parapsychology's most immediate contribution to ethics is on the question of the freedom of the will. This problem is basic; any moral system depends, of course, upon volitional freedom of some degree or other. The more established and orthodox branches of psychology have provided no conclusive answer to this fundamental question of moral relations and have not attempted to do so. The bearing of the psi investigations upon this great issue is therefore all the more important because they are thus far the only source of relevant evidence.

Moreover, the parapsychological discoveries that concern the problem of freedom of the will are very unorthodox. The prevailing psychological view of man today pictures him in physicalistic terms and attempts to integrate him with the physical universe. In fact, any effort to describe mental life as nonphysical in character meets with vigorous opposition by psychologists. Morality and the

institutions depending on it may take for granted, as they must do, that man has a volitional system which exercises free choice of action over his physical environment; but such a view is rejected by all branches of psychology that are dominated by a philosophy derived from the physical sciences.

The psi experiments have shown that human subjects can perform in ways that do not conform to known physical law. Such performances therefore seem unlike manifestations of physical energies; rather, they confirm the traditional hypothesis of a nonphysical mind on which our common-sense concept of free will depends. Obviously, if the entire personality were physical and if it strictly followed the laws of space-time-mass relations characteristic of the physical energies, no true volitional freedom could exist in the individual any more than one wheel of a machine could be free to dominate the machine.

In order to be free, one part of a system has to have properties or powers that are not inescapably subject to regulation by the rest. In showing that the mind does not follow the same principles as the physical body and environment, parapsychology has furnished experimental evidence allowing the inference of real freedom of choice. The mind is not bound in any absolute degree to the physical universe as we know it. In psi demonstrations it gives evidence of being different, nonphysical, or, at least, free from physical law to some degree.

Such freedom is necessarily relative. There can be no interaction of mental and bodily processes without some common properties and principles. The degree of ethical potential of the individual or the species will depend upon a kind of *freedom ratio*, the degree to which the individual mind transcends, or differs in principle from, the physical energetics of the brain. This ratio of course remains to be found. The full scope of the self-regulation possible to mankind depends on the findings. The range of the creative potential of the race will be measured by the extent to which the purely psychical turns out to be unique in nature, governed by its own appropriate laws.

The implied dualism is, as we have said before, only relative. Common underlying properties must be assumed to exist since

inter-operation does occur between the two systems of thought and brain. We cannot yet see very far into this mind-brain relationship; but the mind does function in its own distinctive way to regulate its interaction with matter. That is all the freedom ethics needs, and the case now rests on experimental evidence. It was a necessary beginning.

There is another aspect to the contribution of parapsychology to ethics. We know now that the belief that we may choose our course of conduct is no mere delusion. But the choice itself is determined by our values, especially our appreciation of our fellowmen. The way we treat each other individually or internationally depends on what we think we are.

We respect ourselves the more for knowing we are more than mere machines. We tend to do the same with others. Thus anything that supports the view that men have unique mental powers transcending the material order minimizes those superficial physical differences that divide us and cause discord.

Parapsychology has added evidence to show that behind the space-time-mass relations of a human being are distinct creative functions found nowhere else in nature. Such discoveries make it easier for us to respect the personal worth and aspirations and welfare of the individual. In this matter, too, the psi researches confirm the traditional view of man as set forth in the more advanced religious systems. It was no mere accident that those systems of belief that have most dignified human beings with qualities beyond the physical are the very ones which have also cultured the loftiest conception of fraternal regard among men, the strongest sentiment for peace.

The social binding power of mutual liking and respect and understanding based on nonphysical or spiritual inter-relations is very real and potent. In its growth lies the main hope of human society. Surely the urgent need of our times is to raise the ethical systems which generate this binding power from a foundation of faith to one of factual demonstration. Here parapsychology has begun to make the contribution which the world has most need of, and here it must continue to serve until man's *moral* Renaissance is realized.

J. B. R.

HELP-HINDER COMPARISON IN PK TESTS

By Betty M. Humphrey

ABSTRACT: In the "Help" Section of a series of PK tests, two subjects cooperated in trying mentally to cause six dice to fall with the desired face up. In the "Hinder" Section the same two subjects worked in opposition to one another, each trying to make a different face turn up. The Help Section gave results significantly above chance, the CR being 3.91. In the Hinder Section, the results were slightly positive, but not significantly so. The CR of the difference between the two sections had a probability of .02. When both sections are pooled, the resulting CR is significant, with a probability of .0011. The hypothesis of biased dice is ruled out since the deviations on all faces of the die were positive. The series also gave evidence of the "double decline" position effect, a decline within each half of the 12-trial column on the record page, which has occurred in previous series having the same type of segmented column. The results of this experiment indicate that two subjects working together are more effective than the two working against each other.—Ed.

THE NUMEROUS reports of the PK researches have established the fact that the human mind can directly affect the fall of dice so as to make a specified face turn up more often than would be expected by chance. One of the questions which this fact brings to mind is: What happens if two or more persons try to influence the same dice at the same time? The present research arose as one of the early attempts to answer this question. The experiment was designed to compare the effect of two individuals' working together with that of two working in opposition to each other. The tests were done in a way that permitted comparison between the scores obtained when the two subjects were trying for the same target and the scores resulting when they were trying for two different targets at the same time.

The Plan of the Experiment

This experiment was carried out in April of 1943 at Duke University, three different subjects participating. The two series making up the experiment differ only in the number of throws constituting a "turn" for a subject. In Series A, 36 throws of six dice constituted a turn. This number of trials, however, seemed somewhat monotonous, and it was thought that perhaps the use of

shorter turns might enliven the procedure and serve to keep the subjects' interest at a higher level. Therefore, a turn was shortened to 12 throws. All runs made after this change are designated as Series B.

The six dice used were white ones, $^{11}/_{16}$ of an inch on the edge, with inlaid, tri-colored spots. (The one- and six-faces had black spots, the two- and five-faces had red spots, and the three- and four-faces had blue spots.) The dice were always thrown from a standard dice cup with a corrugated rubber lining. The table upon which they were thrown was six feet by three feet and had the surface and sides padded so that the dice bounced easily.

In each session two subjects participated. The subject who threw the dice while attempting to influence them mentally was designated as the "thrower"; the other subject, who also tried to influence the dice, although he was not throwing them, was called the "observer." These roles were reversed at the end of every 36 throws in Series A and at the end of each 12 throws in Series B. The subject designated as thrower chose his target face for his turn and indicated it to the experimenter (B.M.H.) and to the observer. The observer then determined, by the throw of two dice, whether he was to help the thrower get his designated target or whether he was to try to "hinder" him. If the sum of the two upturned faces on the dice was an even number, he tried to help by his mental attitude; if, on the other hand, the sum of the two upturned faces was an odd number, he attempted to hinder the thrower by trying to make another target face come up more often than the thrower's target. The observer was asked to make his attitude one of just trying to obtain as many hits as possible on his own target, rather than one of actively trying to suppress the thrower's number.

If the throw of the dice indicated that the observer was to hinder, he took as his target the larger of the two numbers uppermost on the dice. In the event, however, that this happened to be the same as the thrower's target, the observer took the number uppermost on the remaining die as his target. The experimenter recorded the respective target faces on the record sheet, which she kept out of sight of the thrower throughout his turn.

After the targets had been decided upon, the throwing began,

with both subjects trying to influence the dice by their mental attitudes. Throughout the turn of the person who was the thrower, the experimenter and the observer took care not to divulge by their speech or by their facial expressions whether the thrower was being helped or hindered. A lively interest in the progress of the scoring was shown by the subjects, and they usually kept close track of their totals. The observer could do this by looking at the record sheet, but the thrower, who was not allowed to see the record, frequently asked for information as to his own total and as to the number of trials remaining.

Rewards of movie tickets and soft drinks were given for high scores, the reward being duplicated for both subjects if they were trying for the same target.

The records were taken on consecutively numbered sheets, each having 24 columns of 12 entries each. (See Figure 1.) For ease in recording, lines after every three entries broke the column into four segments. When the observer was helping the thrower, the total number of hits for each throw of the six dice was recorded in the column, and the total of hits for each segment was placed in the margin beside the column, as shown at the left in the illustration. When the observer was hindering the thrower, the latter's hits were entered in the left-hand side of the column and the observer's hits were recorded on the right. During the test the experimenter made a special effort to mask the noise of the pencil and to vary the recording time so that the thrower would not be able to tell whether one entry or two were being made, and so would not be able to infer whether he was being helped or hindered.

The major problem of the research requires comparison of the results of the data in the Hinder Section with those of the Help Section. The comparison of these sections, together with the evaluation of both sections pooled, is complicated by the negative correlation between the hits of thrower and observer within the Hinder Section. This correlation factor becomes obvious when we consider that a success by the thrower, for example, on any single die throw necessarily means a failure for the observer. The hits for the two participants, when they have different targets, are negatively correlated to the extent of $-.20$. The existence of a correla-

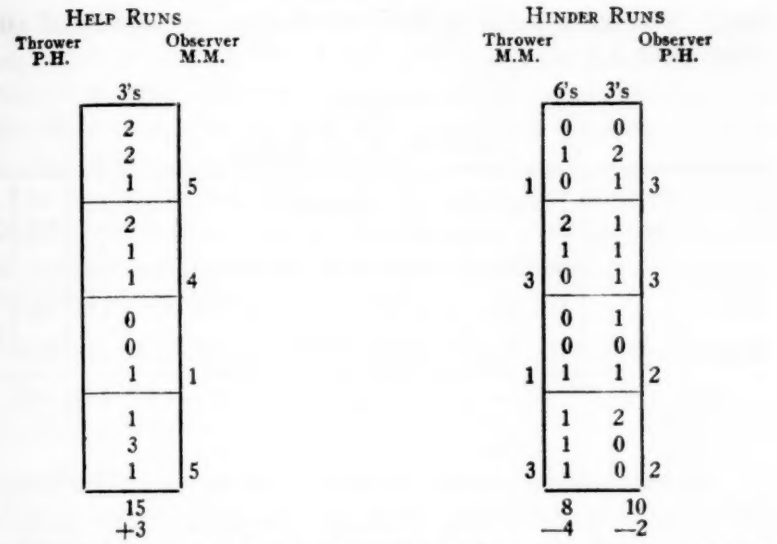

FIG. 1. Sample of a portion of a record sheet for subjects P.H. and M.M. showing the manner of keeping the records for help and hinder runs.

tion between two batches of data requires that the standard deviation be corrected when the data are to be pooled for evaluation. Thus whenever the Hinder Section of the data enters into any evaluation in this report, the standard deviation has been corrected for the correlation factor.[1]

The Help Section is, of course, evaluated by the usual formula since there is no correlation in those data.

RESULTS

Altogether, in this experiment there were the equivalent of 603 standard runs with a deviation of 139 hits above chance expecta-

[1] The conventional formula for the standard deviation of the sum of two batches of correlated data is:

$$\sigma_{\text{sum}} = \sqrt{\sigma_1^2+\sigma_2^2+2r\sigma_1\sigma_2}$$

In the present situation this formula becomes:

$$\sigma_{\text{sum}} = \sqrt{\sigma_1^2+\sigma_2^2+2(-.2)\sigma_1\sigma_2}$$

The SD thus obtained is the SD for the total Hinder Section—the addition of the results of the thrower and observer.

When the Hinder Section is compared or combined with the Help Section, the usual formula for the SD of a sum for uncorrelated data is used:

$$\sigma_{\text{sum}} = \sqrt{\sigma_1^2+\sigma_2^2}$$

In this latter case we would substitute for σ_2 the SD obtained above for the Hinder Section.

tion.[2] This gives a CR of 3.06, which has a probability of .0011. (See Table 1.)

Table 1

SUMMARY OF THE RESULTS

Section	Throwers			Observers			Total			CR	CR_d
	Runs	Dev.	Av. Score	Runs	Dev.	Av. Score	Runs	Dev.	Av. Score		
Help	177	+95	4.54	177	+95	4.54	177	+ 95	4.54	3.91	2.07
Hinder	213	+32	4.15	213	+12	4.06	426	+ 44	4.10	1.31	
Total	.....	.....	.....	.....	.....	.....	603	+139	4.23	3.06	

The main interest in this research, however, lies in the comparison of the Help and Hinder Sections. The Help Section, those runs in which the thrower and observer cooperated, consisted of 177 runs with a positive deviation of 95. This is an average of 4.54 hits per run (4.00 is chance expectation). The CR of 3.91 for this section is highly significant, with a probability of .0001.

The Hinder Section, those runs in which the thrower and observer tried for different faces, consisted actually of 213 runs, but since these are scored both for the thrower's target and for the observer's target, they may be treated as 426 runs. These runs gave a deviation of 44 hits above chance and an average of 4.10 hits per run. The CR of 1.31 for this section is not significant.

If we compare the two sections, we find that the CR of the difference between the two is 2.07, which has a probability of .02.

Position Effects

Because of the irregular alternation between help and hinder runs and because of the change of target for every turn, it is possible to do only one of the customary analyses for position effects. The one type of analysis possible here is the distribution of hits in the column.

For this analysis it seemed best to treat the two series separately

[2] The separate totals for Series A and B which are not pertinent to this discussion are given at the end of the report for anyone who wishes to study the results in detail.

since in Series A the column did not have quite the same importance that it did in Series B. It will be remembered that in Series A 36 throws constituted a turn; thus three columns were involved. In Series B, however, the column was a turn and therefore received more emphasis as a structural unit.

The deviations by segments in the column for the 135 runs in Series A, with help and hinder runs separated, are as follows (the total number of runs and the deviation for the various sections are given under each column):

Help Runs	Hinder Runs			Series Total
	Thrower	Observer	Total	
+ 4	− 2	+4	+ 2	+ 6
+10	0	−5	− 5	+ 5
+ 3	+12	−1	+11	+14
+12	0	−4	− 4	+ 8
45:+29	45:+10	45:−6	90:+ 4	135:+33

Here there are no consistent trends although the total of the hinder runs gives a double decline similar to those found in other series (**1, 2, 3, 4**); that is, there is a decline within each half of the column. Here the effect is obviously due to the canceling of deviations when the throwers' and observers' data are pooled since it is not apparent in the separate presentation of their results. The total of the series shows only slight evidence of the double decline.

The 468 runs of Series B show the double decline much more markedly. Graphic presentation of these results is given in Figure 2, while the numerical data are summarized below:

Help Runs	Hinder Runs			Series Total
	Thrower	Observer	Total	
+32	+ 9	+14	+23	+55
+10	−12	+10	− 2	+ 8
+17	+15	+ 5	+20	+37
+ 7	+10	−11	− 1	+ 6
132:+66	168:+22	168:+18	336:+40	468:+106

When we consider the hinder results of Series B, it can be seen that the double decline effect shows up primarily in the data of the thrower, while a straight vertical decline is found in the work of the observer. The help runs, however, show a very marked double

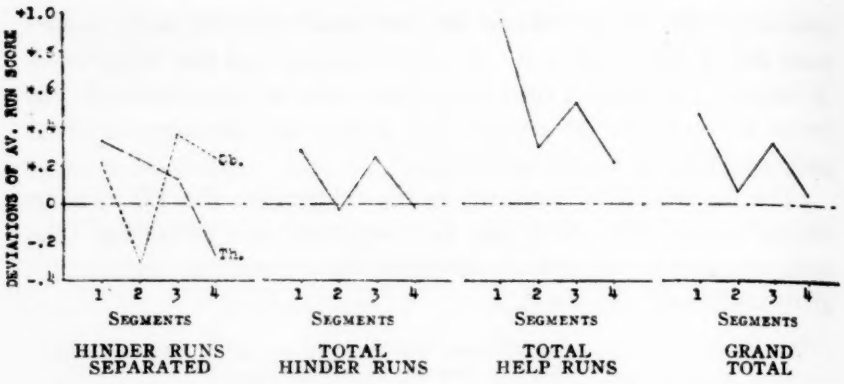

FIG. 2. Vertical distribution of hits by segments in the column for the 468 runs of Series B. (In the first graph the average deviations for the observer and for the thrower are shown separately.)

decline; and of course the pooled total of the two shows the same trend.

When Series A and B are pooled the deviations for the segment in the column are: +61, +13, +51, +14.

DISCUSSION

Before concluding that PK has been operative in producing either the significant results of the Help Section alone or of the total series, we must examine the data to see whether the results may have been due to some other factor, such as faulty dice.

In this experiment all faces of the die served as target at some time during the experiment although not to an equal degree. On the biased-dice hypothesis positive deviations on some faces necessarily have to be counterbalanced by negative deviations on others. Tabulation of the scores for each of the target faces in this experiment shows the following results:

	1's	2's	3's	4's	5's	6's
Runs	69	45	90	117	96	186
Dev.	+16	+10	+21	+31	+15	+45
Av. Run Score	4.23	4.22	4.23	4.26	4.16	4.23

Every one of the target faces has a positive deviation, and the average run scores are remarkably similar. None of them vary much from the average of the entire series, which is 4.23. The

hypothesis of biased dice does not seem, therefore, to be reasonably applicable to these data.

When the approximately 600 runs in these series were completed, it was decided to change to another type of procedure in order to learn more about the PK effect under these conditions. Thus this experiment was considered finished, and a new one was begun under the new procedure. The pooled total of all help-hinder series done up until the time of this writing (September, 1946) is statistically significant and will be reported in separate papers later. Thus it could not be said that this particular series was arbitrarily stopped at a favorable point in an otherwise chance series, especially in view of the fact that the CR of 3.91 can hardly be considered as a marginally significant CR.

The hypothesis that seems to fit these data best is that the subjects were able to exert a definite psychical influence upon the dice—that is, that PK was operative here.

A comparison of the highly significant CR for the Help Section with the insignificant one for the Hinder Section strongly suggests that two subjects working together are more effective than the two working against each other. We cannot determine from this experiment, however, whether the two subjects working together were better than either of the two working alone since none of the subjects had been tested alone.

The observers frequently remarked that they had much more of a feeling of personal challenge in the Hinder Section and that they especially enjoyed the competitive nature of this part of the experiment. Nevertheless, their scores averaged close to chance (4.06). From the viewpoint of the observer it is easy to see how the hinder runs would be for him the most *interesting* since in those he would feel more of a personal responsibility than when he was merely helping the thrower. In spite of this, the observers were not able to obtain much of a deviation. The thrower, of course, was unaffected by these differences of attitude since he did not know whether he was being helped or hindered. It may be that the task of hindering was a distracting one in spite of the fact that the observer enjoyed it more. Also the fact that the observer could see when the thrower was succeeding may have been a hindrance to him.

It is not hard to imagine that from time to time when the observer noted that the thrower had obtained several successive hits he would try harder and harder to obtain hits on his own target. Such an attitude of mental strain, mixed perhaps with a fear of being the loser, may not be at all conducive to good scoring.

In the vertical distribution, the appearance of the decline within each half of the column is of special interest since it has already been found in other series. In the data of the first PK report[3] by Rhine and Rhine (4), in the Hilton, Baer, and Rhine (1) and the Hilton and Rhine (2) researches, this same effect was found in conjunction with the type of segmented record column used in this series. The most recent finding of this effect was in the F. Smith series reported by Rhine (3) in December, 1944.

In each of these researches the column of 12 entries had been broken by horizontal lines after every three entries. This effect has not been found with any other type of recording. The rather surprising thing is that it was not found in the observers' data. The observer could look at the record page, and one would have expected him to be most conscious of the column structure since he could watch the experimenter writing in the segmental totals as each segment was completed. Of course, all participants were well aware of the structure and it is probable that the thrower kept track of his hits mentally by segments also. Evidently the visual aid accessible to the observer was not the important factor in this research in producing double declines. What the important factors in the production of this effect are and just what part consciousness of the structure plays in producing it cannot be determined on the basis of the researches reported.

This experiment is one of the first of a series designed to obtain information concerning the effect of two or more persons' trying mentally to influence the dice. The results of this series raise many problems and suggest several desirable refinements in procedure. For instance, the procedure followed here does not present comparable conditions for thrower and observer; the thrower did not know whether he was being helped or hindered, while the observer

[3] The double decline effect was not found until after this experiment was reported. Therefore it is summarized in the article by Hilton, Baer, and Rhine (1).

knew that he was helping or hindering. In a later series this difference was taken care of by having neither one of the subjects know the target of the other. It is not necessary to review here the other problems suggested by the present experiment since they will become apparent in the publication of the results of the other series.

Appendix

Detailed Summary of Results

	Thrower		Observer		Total		CR
	Runs	Dev.	Runs	Dev.	Runs	Dev.	
Series A							
Help Runs	45	+29	[45	+29]	45	+29	2.37
Hinder Runs	45	+10	45	− 6	90	+ 4	.13
					135	+33	1.57
Series B							
Help Runs	132	+66	[132	+66]	132	+ 66	3.15
Hinder Runs	168	+22	168	+18	336	+ 40	.67
					468	+106	2.60
Both Series Pooled							
Help Runs	177	+95	[177	+95]	177	+ 95	3.91
Hinder Runs	213	+32	213	+12	426	+ 44	1.31
					603	+139	3.06

References

1. Hilton, Jr., H., Baer, G., and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 172-90.
2. Hilton, Jr., H., and Rhine, J. B. A second comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 191-206.
3. Rhine, J. B. The PK effect: early singles tests. *J. Parapsychol.*, 1944, **8**, 287-303.
4. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

Parapsychology Laboratory
Duke University
Durham, North Carolina

A SECOND CLASSROOM ESP EXPERIMENT WITH THE FREE RESPONSE METHOD

By C. E. STUART[1]

ABSTRACT: In a group ESP experiment with free response drawings an experimental psychology class of 14 students produced responses which, when scored directly, yielded only chance results. But a study of displacement showed a significant tendency for the responses to match the next-following stimulus instead of the intended stimulus. Those subjects who reported being rushed or inhibited in performance were compared with the subjects who were satisfied or noncommittal. It was found that the forward displacement observed was wholly attributable to the rushed or inhibited subjects.

Other variables, such as differences of response attitude and differences among stimulus qualities, yielded no significant results.—Ed.

INTRODUCTION

IN AN ARTICLE entitled "A Classroom ESP Experiment with the Free Response Method," published in the JOURNAL for June, 1945, I reported the results of an experiment in which the instructor acted as agent and the members of the class made free response drawings according to certain requested attitudes. For the first two responses they were asked to "concentrate"; for the second two, to make "automatic" responses; for the third two, to "free associate"; and for the last two, to limit their responses to a single idea. The responses were evaluated by an independent judge. The results of that experiment, although not significantly different from chance in total, showed significant variation with respect to the various response attitudes. The "concentrate" and "single idea" responses were significantly better than the "automatic" and "free association" responses, largely because of a significant avoidance of the stimulus picture in the responses under the latter conditions (1).

The present report deals with a repetition of the foregoing experiment. Some changes were introduced into the procedure with

[1] This research was carried out while the writer held the Thomas Welton Stanford Fellowship in Psychical Research at Stanford University. The report has been approved by the Psychical Research Committee and is Communication No. 16 from the Psychical Research Laboratory.

a view to improving it. The experiment was conducted with 14 students of the Psychology 61 Laboratory in Sensation and Perception at Stanford University during the summer quarter of 1943. Professor R. C. Travis and I shared instruction of the class during this quarter. Dr. Travis again served as agent, while I was the experimenter.

Procedure

Instructions

After the students had gathered in the classroom, each was given mimeographed directions for the experiment, 16 blank response sheets, two record blanks, and a piece of carbon paper. I discussed the general nature of the ESP problem briefly and introduced this experiment as one of a number of quantitative approaches to that problem. We went over the instructions in detail to be sure they were clear. They read as follows:

> In this experiment an "agent" will look at a "stimulus drawing" projected on a screen in a room separate from the respondents. He will look at eight stimulus drawings in succession, looking for three minutes at each one with a one-minute pause between presentations. The instructor will time these presentations and tell you the beginning and end of each. He will also remind you before each response of the following attitudes to be assumed.
>
> You are to make your responses on the sheets provided. Use the carbon to make a duplicate of each response. *Write at the top of each sheet your name and the number of the response. Be careful to work alone. Do not be influenced by your neighbor's responses.*
>
> For the responses take the following attitudes as well as you can: Responses 1 and 2: Close your eyes for a moment and visualize in imagination the screen at which the agent is looking. "Concentrate" a moment and try to imagine what is on the screen. If you get any imagery at all try to draw it as your response. If your imagination produces nothing, draw anything at all on the response sheet.
>
> Responses 3 and 4: Imagine momentarily the scene of the agent looking at the screen and then draw anything at all on the response card. "Doodle" as carelessly as you like.
>
> Responses 5 and 6: Visualize momentarily the screen, and then sketch quickly any series of things that come to mind. The task here is to get a train of free associations. If you can't sketch your ideas, write them out.
>
> Responses 7 and 8: Visualize momentarily the screen at which the agent is looking; then try to think of some single concrete object or

idea to draw, and sketch it. Change your mind if you like, but make the response a single object or idea.

Turn in the first copy of your responses.

The next step is to judge whether any recognizable relations exist between the stimulus drawings and your responses. In order to get an unbiased judgment these relations must be inferred before you know the order of the stimuli.

The instructor will show you the eight stimulus drawings one at a time. As each is shown see if among your responses any one occurs *that is more like the stimulus than any of the others.* Use any reason at all for deciding, such as similar idea, similar shape, similar elements, or association of ideas. Then find the *response that matches second best;* then the *third best;* and so on until you have given a rank order to all eight of your responses for the first stimulus drawing. Then rank the eight responses to each of the remaining stimuli in the same way, but *independently* of previous rankings.

Make a copy of your rank orders and hand it in to the instructor.

Stimulus Situation

The stimulus pictures were cut from magazines and mounted on 8" by 5" cards for easy insertion in the balopticon projector. They were selected to provide a group diverse as to topic, form, color, and clarity. The *order* of presentation was randomized by a third person. Tippett's Random Numbers tables were used. The randomization was restrained in such a way as to provide one "complex" and one "simple" stimulus picture in each response attitude. At the time of the experiment no person knew the true order.

The stimulus pictures in the order used were:

1. Flowers. (Hospital room, woman in bed, man greeting her. A large bouquet of flowers is in center background. Colored.)
2. Chick. (Cut-out of chick. Black and white.)
3. Violin. (Cut-out of violin. Colored.)
4. Record. (Greer Garson seated on sofa holding album and two phonograph records. Black and white.)
5. Cigarettes. (Cut-out of package of Old Gold cigarettes. Colored.)
6. Ship. (Two sailors on bridge of ship. Night scene with searchlight prominent. Colored.)
7. Children. (Two children, boy and girl, running in a meadow. Black and white.)
8. Bus. (Cut-out of Greyhound Bus. Colored.)

The agent was in a room about 75 feet distant from the re-

spondents' room. Intervening doors were closed, and time signals were arranged from the respondents' room to the agent by a buzzer system. Upon the prearranged signal the agent took the first stimulus picture from the envelope containing the stimuli, placed it in the projector, and looked at the image thrown on the screen. At the signal indicating the end of the presentation he took the picture out of the projector, selected the second stimulus drawing, and placed it in the projector in readiness for the second presentation. This procedure continued for the eight presentations. The agent kept a list of the order of stimuli. This list he retained until the final check-up. After the eight presentations he shuffled the stimulus pictures in order to break up any cues as to their order of presentation.

Response Situation

I remained in the classroom with the respondents, timed the presentation, and reminded the students of the attitudes to be taken. One of the changes from the previous experiment was a three-minute period for each response instead of the four-minute one used before. While most of the students found this time adequate, a few complained that they felt rushed. At the end of the response period the subjects were asked to report any feeling of ineffectiveness due to being rushed by the timing or being inhibited in performance by the rigidity of the procedure.

Scoring

The respondents handed in the first copies of their responses and retained the carbon copies. I then procured the shuffled stimulus pictures from the agent and set them up, one at a time, for the subjects to rank their eight responses to each in turn. After completion of the ranking the subjects turned in the first copy of their ranking record. I then called the agent to bring in the record of the true order, and, by circling the ranks of the true correspondences, the subjects were able to compute their own scores. The scores used in this report are, however, independently checked from the duplicate unmarked records.

In a study of individual GESP experiments it had been found that the respondent's score differed from an independent judge's

score of the same material (4). Furthermore, in the previous classroom series the responses had been scored by an independent judge. It seemed useful, therefore, to have the present responses independently scored as well.

The eight responses of each subject were separated into two sets of four, with the appropriate stimuli divided likewise. This separation was necessary because of the change in form of the response with different attitude instructions. The attitudes "concentrate" and "draw single idea" produce relatively well organized drawings. The attitudes "automatize" and "free associate" produce relatively unstructured responses. If these systematically different kinds of responses confront the judge repeatedly he may match the simple stimuli with the organized responses and the complex stimuli with the unstructured responses on the grounds that these are points of similarity to justify ranking. His scores would then be intercorrelated and open to the "stacking effect" hypothesis. (See Discussion section.) By separating the responses into two sets, as was done here, the judge does not have to rank two different kinds of responses in the same matching task.

Set 1 consisted of Responses 1, 2, 7, and 8. Set 2 consisted of Responses 3, 4, 5, and 6. Each of these sets was preferentially matched by two judges in the customary way. To match four items the judge (unaware of the true order of either stimulus pictures or responses) lays the four stimuli before him in a row. He takes one of the responses and decides which of the four stimuli is most like the given response in form, meaning, or other criterion of association. He then decides which is the next most likely association, the third most likely, and the poorest match. He records these ranks for the first response. He then ranks independently each of the other three responses in the same way. After this first matching the four responses are laid before him. He now independently ranks these four responses with respect to each of the *stimuli*. A set of stimuli and responses and the method of scoring are illustrated in a previous article (2). The independent judge's score is the combined rating of two judges, Miss Elizabeth A. McMahan and Mrs. Elizabeth Dupuy, both of the staff of the Parapsychology Laboratory at Duke University. Two judges were em-

ployed in order to get a score distribution strictly comparable to that of the majority of other free response work now reported.

The results thus include ranking of the responses by the subjects on a 1 to 8 scale, as described in the instructions, and ranking by independent judges on the customary 1 to 4 scale. In each case both direct and displaced correspondences are scored. These four score distributions have different expected means and standard deviations. Since nothing is served by retaining this arithmetical variety, all scores have been transformed to "standard" scores. The standard score is the deviation of the observed score from its theoretical mean chance expectation, divided by the standard deviation of the observed score distribution. The sign of the standard score is positive if the original deviation is in the direction of better-than-chance correspondence between the matched items. It is negative if the deviation is in the direction of a poorer-than-chance correspondence.

For example, subject *Gir* ranked his direct correspondences with ranks 5, 1, 3, 5, 2, 2, 3, 2 on a 1 to 8 scale. These ranks total 23. Since by chance all ranks are equally likely, the mean rank is $(1 + 8)/2$, or 4.5, and the mean chance score is 8 times the mean rank, or 36. The deviation is 13. The standard deviation of all the subjects' total rank scores is 6.94. The deviation divided by the standard deviation gives a standard score of 1.9. This score would have a positive sign, since the low rank score is in the direction of a better-than-chance correspondence.

The standard score is a form of critical ratio and may be generally interpreted like other CR's. But in the present report there are only 14 subjects to provide the score distributions, and considerable correction would be necessary to make the standard scores strictly comparable to large-sample CR's. For this reason, and other reasons cited in previous reports of the free response method, the "Student" *t*-test is applied here for tests of significance.

Results

The results are summarized by subjects in Table 1. After a designated subject the letter *i* indicates his report of being rushed or inhibited; the others were noncommittal. Direct and displacement scores are given. Under "Subject" the score of the subject's own ranking is given. Under "Ind." the score of the independent

judges' rankings is given. Under "Both" is given a total score derived from pooling the rankings of both subjects and independent judges. The *t*'s at the bottom are calculated from the mean deviations from chance of the raw data (not the standard scores). The probabilities cited are for *t* with 13 degrees of freedom.

Table 1

DIRECT AND DISPLACEMENT STANDARD SCORES

Subject	Direct			Displacement					
				Backward			Forward		
	Subj.	Ind.	Both	Subj.	Ind.	Both	Subj.	Ind.	Both
Cri i.....	−1.0	− .7	−1.0	+ .7	− .3	+ .2	+ .8	+1.1	+1.0
Fli i.....	+ .9	+ .8	+1.0	+ .1	−1.4	−1.2	+ .6	+ .6	+ .7
Gri	+1.9	+ .3	+1.5	− .7	+1.9	+1.2	−2.0	+ .2	−1.1
Ing	− .6	+1.0	+ .2	+ .3	+ .7	+ .8	− .5	−1.8	−1.3
Kuh	+ .7	+ .5	+ .8	− .9	− .6	−1.1	+ .1	+ .2	+ .1
Low	− .1	−2.8	−1.7	−1.0	+ .3	− .4	+ .2	+ .2	+ .2
Naf i.....	− .1	−1.3	− .8	+ .3	+ .9	+ .9	+1.3	+1.5	+1.6
New i.....	− .4	+ .2	− .2	+2.0	+ .6	+1.8	+ .9	+2.6	+1.9
Pat	+ .3	−1.0	− .4	+ .3	+ .6	+ .7	+ .3	+ .6	+ .5
Pei i.....	− .7	− .7	− .8	−1.2	− .4	−1.2	+2.1	+ .3	+1.4
Ray i.....	+ .7	+1.1	+1.1	− .5	−1.0	−1.2	+1.0	+ .3	+ .8
Rob	−1.4	− .3	−1.1	+1.4	−1.0	+ .1	+1.6	+1.7	+1.8
San i.....	−1.3	− .5	−1.1	− .7	+2.0	+1.3	+1.4	+1.1	+1.4
Sym i.....	−1.4	− .7	−1.2	+ .5	+ .6	+ .8	+1.6	+1.7	+1.8
t.....	− .71	−1.00	−1.02	+ .15	+ .70	+ .70	+2.44	+2.59	+2.84
Prob.....	.245	.168	.162	.440	.248	.248	.014	.011	.007

When the responses were ranked with respect to the intended stimuli—the direct correspondences—scores were slightly negative, but not significantly different from chance. No consistent relationship between the stimuli and responses can be inferred from the total *t* of 1.02. The subjects' scores and the independent scores give virtually the same total result but show little correlation.

In the previous classroom experiment the limited response attitudes ("concentrate" and "single idea") yielded results significantly different from the unlimited response attitudes ("automatize" and "free associate"), the latter attitudes giving significantly negative scores. In the present results this observation is not borne out. The score for the limited response attitudes is −1.52, for the un-

limited attitudes, +.22. The direction of difference is reversed and the difference is not significant. Examination of the independent scores alone (to make the situation more comparable to the earlier experiment) gives scores of −.60 and −.26, respectively, a negligible difference.

The variation among the four response attitudes is likewise not significant. Scores for the four attitudes in order are: −.58, +.04, +.27, and −1.61. A chi-square test shows this variability to be well within chance limits.

The stimuli were paired in such a way that under each attitude condition there would be one cut-out stimulus picture in which a single form was emphasized and one "complex" stimulus picture in which there were a number of items that might be attended to. Comparison of the effectiveness of these two categories of stimulus shows for the simple stimuli a score of −1.12; for the complex stimuli, a score of −.18. The difference is not significant.

Nor was there evidence of any individual stimulus being significantly more effective than the others. The scores for the eight stimuli are, respectively: −.44, −.38, +.32, +.25, −1.46, +1.84, −1.52, −.76. A chi-square test shows that the observed variation has a chance probability of .27.

I had asked the subjects, after completion of their responses, to report any feeling of ineffectiveness due to being rushed by the timing of the responses or inhibited in performance by the framework of the experiment. Four subjects reported being rushed; and four others said they felt inhibited in expression. This left six subjects who were either noncommittal or generally satisfied with the experimental conditions. For the direct correspondences the score for the rushed and inhibited subjects is −1.14; the score for the satisfied or noncommittal subjects is −.23. This difference, while in the direction of negative performance by the dissatisfied subjects, is not significant.

Displacement

The responses ranked with respect to the just-preceding stimuli are negligibly different from chance, with $t = .70$. There is thus no evidence of a consistent backward displacement as was found in clairvoyance studies (2).

The responses ranked with respect to the just-following stimuli appeared to show a consistent tendency. The scores of 12 of the 14 subjects were positive. Subjects' scores and independent scores are in general agreement and have a correlation of +.67. The total *t* of 2.84 is significant, with a P of .007. The tendency appears to be that of a forward displacement.

The significance of this last result may immediately be questioned on the grounds of selection of one of four possibilities (positive or negative deviation of the direct scores and positive deviations of backward and forward displacement). Multiplication of the P-value by 4 leaves a probability of .03, above the one per cent criterion of significance. But a further observation tends to substantiate the validity of the forward displacement.

As noted above, the rushed and inhibited subjects had shown only slightly more negative scoring than the satisfied subjects. The forward displacement scores, however, are markedly different. (See Table 1.) The score of the rushed and inhibited subjects is +3.52, and the score of the satisfied subjects, +.03. These have a *t* of 6.56 (P = .0001) and a *t* of .03, respectively. The *t* of the difference is 2.66, which, with 12 degrees of freedom, has a probability of .01. It is evident that all the displacement may be attributed to the eight subjects who reported being rushed or inhibited.

Discussion

Since the forward displacement is the only valid finding here, consideration of counterhypotheses to ESP may be limited to that part of the data. The experimental conditions were designed to exclude sensory or rational cues as to the order of the stimuli. Randomization was accomplished with the use of random numbers except for the order of presentation of the stimuli to the subjects for their ranking. In the latter case the agent shuffled them, and then they were shuffled again by the experimenter. First copies of the responses and the subject's ranking record were turned over to the experimenter before the order of stimuli was known to anyone but the agent. Hence no record could be altered. All data and computations have been double-checked to exclude error.

There remains, however, a counterhypothesis intrinsic to the

method that must be considered: the "stacking effect" that may occur when a number of responses are made to single stimulus pictures. The scores are evaluated by a *t*-statistic, which assumes that a given subject's response is independent of the responses of the other subjects. If one subject were to copy the drawings of another, or if a chance event, like the sounding of a fire siren, should suggest to many of the subjects the topic of a fire for their responses, the independence assumption would not be fulfilled. The chance mean would not be affected, but the true variance would be greater than the variance estimated from the subject's scores, and the *t* would be spuriously high.

An analysis of variance of the actual ranks given by the subjects to the forward displacement correspondences should provide evidence of any excessive trial-to-trial variance. Completely independent responses and rankings both between subjects and among their responses should yield a variance per stimulus-response rank of 5.25 (3). An analysis of variance of the subjects' rankings provides four estimates of variance from the raw data. The estimate from all the ranks together is 5.71. The estimate from the variation among subject total scores is 8.19. The estimate from the variation among the total scores for each trial of the seven stimulus-response correspondences is 3.77. The estimate from other uncontrolled sources of variation is 5.45. None of these estimates of variance are significantly different from one another or from the hypothetical variance. If the stacking effect were of serious concern in this experiment the variance estimated from trial-to-trial variation would have been large. It is, on the contrary, rather small. As would be expected, a *t*-test of the significance of the forward displacement, based upon the means of the trials alone and thus excluding any stacking effect, yields $t = 3.59$, which, for six degrees of freedom, gives $P = .006$.

The positive bearing of this experiment is limited, but perhaps not unimportant. In clairvoyance experiments the phenomenon of consistent *backward* displacement was noted. Two obvious guesses at an explanation were: (1) that the displacement was due to a removal of the frustration character of the previous stimulus after

it had been dealt with; and (2) that there was characteristically a time lag or latency for the ESP impression, and the subject, by making a free rather than a studied response, kept one step ahead of the ESP impression.

The observation of consistent *forward* displacement would disfavor the latency hypothesis and suggest pursuit of a psychological rather than a physical explanation.

The agent had the manila envelope enclosing the stimulus pictures yet to be presented on the table beside the projector. There is no easy explanation of why the top picture in the envelope should have become more important than the brilliantly illuminated picture in the projector, or its projection on the screen with which the attention of the agent was occupied.

The second phenomenon, that displacement occurred significantly in the responses of the rushed and inhibited subjects, emphasizes the psychological character of the displacement but again offers no suggestion as to why it was consistently one step ahead. The subjects were not told about the displacement hypothesis and would not therefore consciously adopt displacement as a substitute task or goal. The orderly nature of the result would suggest that something in the stimulus situation determines the kind and direction of displacement.

The failure of the experiment to show a difference under limited and unlimited response attitudes, as found in the former experiment, is of uncertain bearing, as are all of the other comparisons giving chance results. In early work with the free response method there seemed promise that the drawing type of response could be recognized as having a direct relation to the stimulus or virtually none at all. Thus valid differences in conditions would show up in even a brief study with a few subjects. Extended use of the method has shown that, on the contrary, other things being equal, results with drawings are not greatly different from those with card-guessing tests. Any but sporadic instances of differences in stimulus effectiveness or response attitudes would, accordingly, require planning of a much greater number of observations than these exploratory studies have utilized.

REFERENCES

1. STUART, C. E. A classroom ESP experiment with the free response method. *J. Parapsychol.*, 1945, **9**, 92-105.
2. ———. An ESP experiment with enclosed drawings. *J. Parapsychol.*, 1945, **9**, 278-95.
3. ———. An ESP test with drawings. *J. Parapsychol.*, 1942, **6**, 20-43.
4. ———. GESP experiments with the free response method. *J. Parapsychol.*, 1946, **10**, 21-35.

Parapsychology Laboratory
Duke University
Durham, North Carolina

TARGET PREFERENCE IN PK TESTS WITH DICE

By J. G. Pratt

ABSTRACT: The extensive records of a successful PK subject, L.H.G., in the Gibson Cup Series show average scoring rates for the six faces which are directly related to their numerical order. This raises the problem whether target preference was a factor which affected this subject's PK performance. The answer was sought in a comparison of the position effects found within the set when the 1-2-3 faces were the targets and when the 4-5-6 faces were used. More pronounced position effects were found when the latter three faces were the targets. Not only were these hit patterning effects highly significant in themselves, but they were significantly dissimilar from the position effects found for the 1-2-3 faces. Evidence was found that the use of the five-face, which was most preferred by the subject, produced the most pronounced scoring trends. No hypothesis other than differences in the subject's psychological reactions to the targets seems capable of explaining the findings. The target preference appears from a chronological analysis of the series to have developed as the subject gained experience in the experimental situation.—Ed.

Introduction

In many of the PK experiments with dice significantly higher scores on some faces than on others have been observed. When the dice used have been of the type which have the faces marked by small painted depressions, the use of the higher faces as targets has commonly given the higher scores (**1, 2, 3, 5, 6, 7**). Perhaps the most obvious hypothesis suggested by this finding is that the dice were biased, and the investigators have dealt fully with this question, along with the various other counterhypotheses, before reaching the conclusion that PK occurred.

As the research advanced to the point of being concerned with problems of the nature of PK, the question arose: Do the differences in scoring on the various targets require, at least in part, a psychological explanation? Does the subject's psychological attitude toward the different faces affect his PK performance?

This problem is one which has seemed difficult of solution experimentally because of the absence of a method for distinguishing between psychological and physical factors in the production of dif-

ferent score averages for different targets. However, the study of position effects in PK test records opens up one avenue by which a solution may be reached.

What is the logic of this approach? It recognizes that there is nonrandom patterning of successes in the test records which can be explained only on the PK hypothesis. These significant position effects themselves, then, are a special kind of evidence for the occurrence of PK, and significant variations in position effects from one part or condition of the test to another may reflect differences in PK performance. If data which are known to show significant position effects are grouped according to the different targets used, and if this leads to the discovery that the target groups give hit patterning effects which are significantly dissimilar, an explanation in terms of differences in the subject's psychological reactions to the targets would be indicated.

Specifically, this study deals with the problem of whether there is any evidence of target preference in the records of subject L.H.G. in the Gibson Cup PK Series. On all counts, these data are particularly appropriate for this investigation. The series is a large one in terms of number of total trials, so that it is safe to generalize on the basis of various subdivisions of the whole. These records show position effects which are both highly significant statistically and remarkably lawful in the consistency of their occurrence throughout the series (**4**). Finally, all faces of the dice (which were of the common commercial variety with excavated spots) were used as targets, and the averages obtained varied directly with the denomination of the faces.

It is necessary, before proceeding with the present investigation, to review two experimental reports which have presented specific evidence that the favoring of the higher faces involved mental factors. Both series were the work of a single subject, M.P.R., and each involved the throwing of a pair of dice with a combination of faces as the target. One was a comparison between high dice and low dice (**5**), and the other, between double sixes and double ones (**6**).

In the former study the High-Dice Series averaged 5.59 hits per run (where 5.00 is the expected chance average) and the Low-

Dice Series averaged 5.21. The difference was significant. But when the comparison was restricted to the salient first three trials on the page for each target, the advantage was the other way around, with the Low-Dice Series averaging 7.27 as compared with 6.77 for the High-Dice Series.[1]

The other study, in which double sixes and double ones were the targets, similarly showed an over-all favoring of the higher faces. The two targets averaged 1.23 and 1.09, respectively, where 1.00 is the expected chance level. When each set of target records was checked for the opposite double as a control, both double sixes and double ones were found to have occurred with less than the expected frequency and in about equal proportion. Furthermore, the advantage in favor of double sixes in the experimental series was found to be restricted to the first quarter of the calendar period of the tests.

The results of these two experiments were interpreted as indicating, but not as proving conclusively, that the difference between the targets was psychological rather than physical. It is not difficult to conceive of the subject's attitudes toward the two targets in each experiment as shifting with respect to one another as the testing progressed through the page or series, but it is hardly reasonable to suppose that the dice were biased in one part of the page or experiment and not in another.

Data and Procedure for the Present Investigation

The last number of the Journal contained a report of an intensive study of position effects in L.H.G.'s work in the Gibson Cup Series (4). Only a brief statement of the conditions of the original experiment and the form in which the records were kept is necessary here, as a full description of the series was given in that report. Six dice were thrown at a time from a cardboard cup onto a blanketed surface while the subject "willed" them to come to rest with a specified target face uppermost. The number of hits were observed immediately after each throw, and these were recorded for

[1] The averages originally reported were 7.14 for high dice and 6.67 for low dice. The slight discrepancy arises from the fact that Reeves and Rhine included in their study three pages of sets done by incidental subjects, while the analyses upon which these figures are based were restricted to the work of the major subject, M.P.R.

six successive throws in a short column of six entries. The total of the six trials was entered directly below the column. The same target was used throughout 60 throws, or 10 parallel columns, and the score of all 10 columns was then computed and recorded. As many as eight such sets were recorded on a single page, with four as the usual number on a witnessed page and six on an unwitnessed page.

One large section of L.H.G.'s work, which was witnessed and recorded by E.P.G., the experimenter, gave the following averages of hits per 24 dice for the faces taken in ascending numerical order: 4.00, 4.05, 4.00, 4.12, 4.25, and 4.36. In another section in which L.H.G. was unwitnessed the corresponding figures were: 4.19, 4.10, 4.12, 4.15, 4.59, and 4.69. Both sections of the data showed a remarkable regularity of increase in average from the one- to the six-face.

The method for investigating whether a psychological factor entered into the production of these averages in the present study was that of dividing all the sets into two groups: those in which the 1-2-3 faces were the targets, and those in which the 4-5-6 faces were used. It was also decided to keep the witnessed and the unwitnessed records separate because of the striking differences in position effects between the two sections of the data. For convenience, the two parts of the series will be designated hereinafter as the W-Section and the U-Section, respectively.

In only five sets, all of which were from the W-Section, did the subject change targets in the middle of the set, and these have all been omitted from this study of target preference.

Results

Position Effects for the Two Target Groups

Since the U-Section was found to show the more pronounced position effects, the sets from that section were selected as the first to be studied by dividing them into the two groups, one for the 1-2-3 faces and the other for the 4-5-6 faces. The results of this separation are shown in Figure 1. As in the report in the last issue, the rate of scoring for each of the 60 trial positions in the composite set is shown in terms of average number of hits per 24

dice. This makes it possible to trace any difference in position effects which seems to be attributable to target preference right through the various columns of the set.

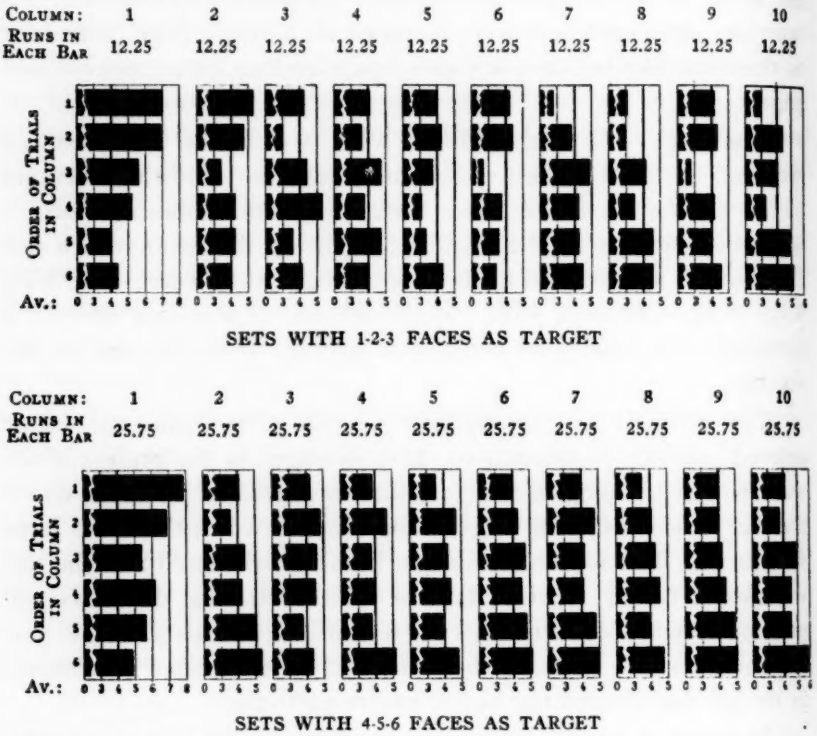
COLUMN: 1 2 3 4 5 6 7 8 9 10
RUNS IN EACH BAR 12.25 12.25 12.25 12.25 12.25 12.25 12.25 12.25 12.25 12.25
ORDER OF TRIALS IN COLUMN
AV.:
SETS WITH 1-2-3 FACES AS TARGET
COLUMN: 1 2 3 4 5 6 7 8 9 10
RUNS IN EACH BAR 25.75 25.75 25.75 25.75 25.75 25.75 25.75 25.75 25.75 25.75
ORDER OF TRIALS IN COLUMN
AV.:
SETS WITH 4-5-6 FACES AS TARGET

FIG. 1. Composite sets for all U-Section data separated into two target groups.

The fact that the total scores for the 1-2-3 faces were below those obtained from the 4-5-6 faces is clearly reflected in Figure 1. The over-all average of the composite set for the former group, shown in the top half of the figure, is 4.16, while that for the 4-5-6 faces is 4.57. With 735 runs in the former group and 1,545 in the latter, the difference in average, .41, is highly significant (SD_d = .082; CR_d =5.02).

This difference, however, is not reflected in the first half of the first column of the set, where the highest scoring occurred. While the first trial of the set gave a higher average for the 4-5-6 faces

(7.84 as compared with 7.02 for the low-scoring faces), the second and third trials actually yielded a higher average score on the 1-2-3 faces. This general comparison of the two target groups in the first trials of the set, where PK was at its best, may reasonably be based upon the combined first three trials. No further justification for selecting these three trials is required, as the comparison is not essential to any conclusion reached in this study. In these three positions as a group, the 1-2-3 faces average 6.53 and the 4-5-6 faces, 6.71. The difference, .18, is insignificant, and the two figures are remarkably close for two results which show as much deviation from expectation as these two do. *Thus both groups of targets were practically on a par at the start of the set, and the significantly higher scoring on the 4-5-6 faces developed as the throwing proceeded into the body of the set.* This result agrees with that obtained for the page in the M.P.R. comparison of high dice and low dice as targets. There, it will be recalled, the two targets were about equal in the first three trials—actually, the low-dice average was slightly better—and the over-all higher average for high dice was obtained in the remainder of the page. In the U-Section of the present series, therefore, as in the M.P.R. study, there occurred for the 1-2-3 faces as targets a greater decline in average scoring rate after the first few trials than for the 4-5-6 faces. This confirmatory fact from the present series greatly strengthens the evidence from earlier findings, though it alone still does not conclusively prove that the difference in rate of scoring for the different faces of the die must be partly explained in psychological terms.

Fortunately, it is possible to carry this investigation even further in the L.H.G. series and to find out how the statistically significant difference between the two groups of targets within the sets actually came about in terms of differences in hit patterning or position effects. It is important to recall here that one significant feature of the U-Section data was the incline in scoring in the pooled eight middle columns of the set. The combined trials in the first position of these eight columns averaged 3.91; those in the last position, 4.70. Each average is based upon 304 runs, and the CR of the difference is 5.34.

Inspection of these same columns in the two parts of Figure 1

is sufficient to show that this incline in scoring occurred chiefly in the sets in which the 4-5-6 faces were the targets. The statistical evaluation of the preceding paragraph (all *first* trials vs. all *last* trials of the eight middle columns) may be applied to the composite set of each target group as a first rough approximation of this difference in position effects. For the 1-2-3 faces, with a total of 98 runs in each combined trial position, the beginning throws of the middle columns average 3.98 as compared with 4.19 for the ending throws, a difference of .21, which has a CR of .82. The 4-5-6 faces, with 206 runs in each pooled position, average 3.87 for the beginning throws and 4.94 for the ending trials. The difference of 1.07 in average has a CR of 5.94. It is clear that the position effects of these columns is highly significant when the 4-5-6 faces were the targets and that it does not reach the level of significance in the records based upon the 1-2-3 faces. The question of whether the difference between these two results is itself a statistically reliable one will be considered later.

At this point in the target preference study consideration was given on purely logical grounds to what one would expect if the data of the W-Section were analyzed according to target groups in the same manner. In the preceding report on position effects in L.H.G.'s results it was found that the middle columns of the set in the W-Section showed both beginning and ending salience, whereas these columns in the U-Section showed ending salience only. Since the middle-column *ending* salience of the U-Section sets occurred largely with the use of the 4-5-6 faces as targets, it seemed logical to expect that these same faces would account for the middle-column *double* salience of the W-Section sets. The prediction was made that the curve of terminal salience in these data, as measured by the difference between the combined two end positions vs. the four central ones, would be more pronounced for the 4-5-6 faces than for the 1-2-3 faces.

The results of dividing L.H.G.'s sets of the W-Section into the two target groups are shown in Figure 2. Because of the fact that the position effects of the middle columns are less pronounced here than in the sets of the U-Section, the abbreviated form of the set, in which the middle columns are pooled, has been adopted.

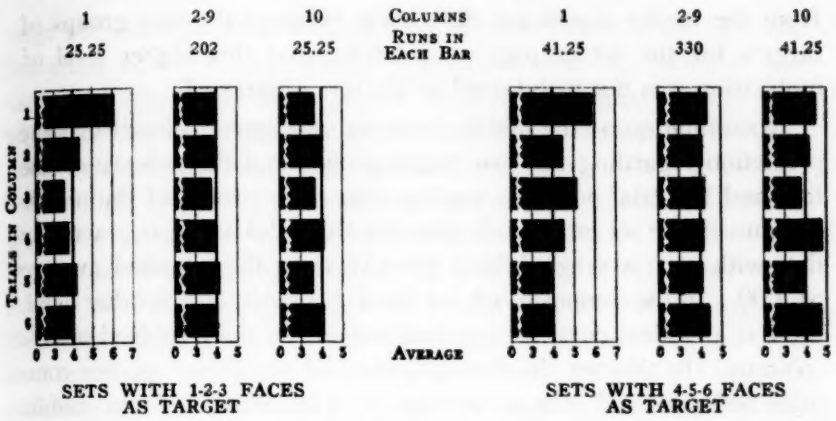

Fig. 2. Composite sets for all W-Section data separated into two target groups.

Here, again, the first point of interest is the comparison of the rates of scoring of the two target groups at the beginning of the set. The W-Section as a whole showed a remarkably high rate of scoring only in the very first trial of the first column, so the comparison may be based on this single trial in each of the two target groups. Figure 2 shows that, in this first trial, the 1-2-3 faces as targets gave a higher average than that obtained with the 4-5-6 faces (6.10 as compared with 5.89). In contrast to this, the total average of all W-Section results for the 1-2-3 faces was 4.01, and that of the higher faces, 4.28. There are 1,515 runs based on the 1-2-3 faces and 2,475 runs in the other target group; and this difference of .27 in over-all average is highly significant, as the CR_d is 4.55. Again, the observation first made in the M.P.R High-Dice Series and Low-Dice Series is confirmed in these records, for the higher faces are found not to have any advantage in the first throw of the set, where the conditions were most favorable for the occurrence of PK.

This line of evidence does not prove, of course, that the dice were not biased, for it is possible that the subjects were more strongly motivated at the start of the set or page to succeed with the more difficult faces. There may have occurred at that point in the tests a stronger motivation when throwing for the 1-2-3 faces as a result of the challenge presented by them. If so, it is evident,

from the highly significant differences between the two groups of targets for the set or page as a whole, that this higher level of motivation was not maintained as the test progressed.

Examination of the middle columns of Figure 2 shows that the prediction regarding the two target groups and the salience of the first and last trial positions was correct. The profile of the middle columns in the set on the left, that for the 1-2-3 faces, is practically flat, with each average falling quite close to the expected average of 4.00. In the composite set for the 4-5-6 faces, on the other hand, there is a distinct curve of terminal salience in the combined middle columns. In this set the first and the last positions together comprise 660 runs and give an average of 4.38, while the four middle positions, with 1,320 runs, average 4.15. The average difference, .23, has an SD of .087 and a CR of 2.64, which is significant at the one per cent level ($P = .004$). In the body of the set for the W-Section, therefore, as for the U-Section data, we find that significant position effects occurred only when the subject was using the 4-5-6 faces of the dice as her PK target. These findings in themselves, even without further efforts at evaluation, point to the conclusion that the subject's psychological reactions to the different faces of the dice actually affected her PK control over them.

Evaluation of Differences in Position Effects for the Two Target Groups

In both the U-Section and the W-Section, the position effects in the middle columns of the sets were significant for the 4-5-6 faces as targets, but not for the 1-2-3 faces. It was desirable to find out whether the differences in position effects were themselves statistically significant. Before this step could be taken, it was necessary to express the position effects in numerical form.

In the case of the U-Section results, the observed difference in the scores made in two successive trial positions, the end of one column and the beginning of the next column within the set, provided a measure of position effect. There were nine observations for each set. The decision to use the change in scoring from the end of one column to the beginning of the next as the basis of this test was made without knowledge of how it would compare with

other possible measures, such as the difference between the first and last trials in the same column. The choice had two practical advantages: these trials were easier to compare because they were adjacent ones in the records as they had been transcribed for purposes of these analyses; and extending the observations to include the last trial of the first column and the first trial of the last column increased the number of observations by one-eighth.

The expected average chance score, with six dice thrown at a time, is one hit per throw. Inasmuch as a drop in scoring in going from the end of one column to the beginning of the next had been established as the general pattern, the actual difference found in each case was recorded as a positive value when it conformed to this pattern and as a negative quantity when it did not. When the scores in the two trial positions compared were equal, a zero was recorded.

In this manner, two populations of observed frequencies for contiguous ending-beginning trial differences within the sets of the U-Section were obtained, one for the 1-2-3 faces and the other for the 4-5-6 faces. The results of the statistical evaluation of the difference between these two groups, as shown in Table 1, indicate that the dissimilarity is of a degree which would be equaled or exceeded by chance an average of 14 times in 1,000 such comparisons. This probability is below the one per cent standard of significance.

The results of the W-Section, with salience in both the beginning and the ending positions of the middle columns, presented a

Table 1

STATISTICAL EVALUATION OF THE POSITION EFFECT DIFFERENCE BETWEEN THE TWO TARGET GROUPS IN THE U-SECTION

	1-2-3 Faces as Targets	4-5-6 Faces as Targets
Number of Observations	441	927
Mean of Distribution	.0884	.2567
Sigma of Distribution	1.3142	1.3642
Sigma of Mean	.0626	.0448

Difference between Means	.1683
Sigma of Difference	.0768
Critical Ratio of Difference	2.19
Probability of Random Occurrence	.014

somewhat different problem regarding what data to use for a statistical study of the difference between the two target groups. It was decided, again without any knowledge of the result to be expected, to use as the basic datum the observed difference between each end position of each middle column and the average of the adjacent two trials in the same column. Thus the first position score would be compared with the average of the second and third combined, and the sixth position score, with the average of the fourth and fifth. When the hits in an end position exceeded the average number for the two adjacent trials, the difference fitted into the established pattern and it was recorded with a positive sign. Differences in the opposite direction were recorded as negative quantities, and equal cases were recorded as zeros. The statistical evaluation of the difference between the two frequency distributions obtained by this procedure for the two target groups is summarized in Table 2. The CR of the difference is 1.86. P = .032, which is not independently significant.

Table 2

STATISTICAL EVALUATION OF THE POSITION EFFECT DIFFERENCE BETWEEN THE TWO TARGET GROUPS IN THE W-SECTION

	1-2-3 Faces as Targets	4-5-6 Faces as Targets
Number of Observations	1632	2608
Mean of Distribution	− .0021	.0560
Sigma of Distribution	1.1018	1.1533
Sigma of Mean	.0215	.0226

Differences between Means	.0581
Sigma of Difference	.0312
Critical Ratio of Difference	1.86
Probability of Random Occurrence	.032

The results of Tables 1 and 2 are statistically independent, so that it is quite proper to test their joint significance by the method of squaring the critical ratios and treating their sum as chi squared. This gives for the U-Section and W-Section data combined a chi-square value of 8.26. With two degrees of freedom, P = .02. This test, however, ignores the fact that both critical ratios are alike in sign; that is, that the same targets, the 4-5-6 faces, gave

the more pronounced position effects in the sets of both sections. When the above probability is corrected to allow for the fact that the analysis of the W-Section results came out, as predicted, to agree with the findings from the U-Section, the difference in effect between the 1-2-3 faces and the 4-5-6 faces as targets is significant at the one per cent level.

Target Differences in Relation to Other Criteria of Face Preference

Up to this point the study of target preference in L.H.G.'s results has proceeded on the assumption that any difference in L.H.G.'s reaction to the various faces would show up in the contrast between the two groups of targets. This procedure has involved the hypothesis that the difference in average for the two groups of faces was due, in some degree at least, to psychological or PK factors. It is perfectly possible, however, that imperfections in the dice were the predominant factor in causing the differences in total score averages among the six faces. The subject could have preferred one of the lower faces, and this might have resulted in an advantage to the PK function in the use of this face as the target. However, if the favoring psychological factor were weaker than the opposed physical one of unbalanced dice, the averages might still have been directly related to the numerical value of the faces. This, in itself, would not have prevented the discovery of an effect of target preference through an analysis of differences in position effects. However, it is not unreasonable to suppose that the subject might have had a preference for *two* faces. If this were the case, and if one of these happened to be in the 1-2-3 target group while the other was in the 4-5-6 group, the division of the records on the basis of target face averages might not have revealed the factor in question.

These possibilities suggest the importance of using criteria of target preference which are more definitely indicative of some such psychological factor. One such criterion is what the subject says about her preferences for the different faces throughout the series. Another is the frequency with which the subject elects to throw for the different faces when she is entirely free to follow her own bent.

As no written record was made of the subject's expressed pref-

erences at the time of the tests, the experimenter and the subject were asked by correspondence to state independently which, if any, of the six faces L.H.G. preferred. Their answers indicated unequivocally that she stated during the period of the experimentation that she liked the five face best.

We can be certain that the choice of targets was made by the subject only in those sets in which she was working alone; hence only in the U-Section data is the relative frequency of the different targets an indication of her face preferences. In that part of the records, the five-face was the target for 52 sets, the four and the six together were used for 51 sets, and one of the other three targets was used in each of 49 sets. We may therefore divide the U-Section sets into these three target groups in order to test the assumption that we have reflected here three different degrees of preference for the faces. If this assumption is correct, the five-face should give the most pronounced incline in scoring in the middle columns of the set, the four-face and the six-face combined should be intermediate, and the 1-2-3 faces as targets should show the least effect.

The results shown in Table 3 support this hypothesis of degrees of preference as reflected in frequency of choice of targets. The

Table 3

DIFFERENCE BETWEEN FIRST AND LAST TRIALS OF MIDDLE COLUMNS OF SETS IN U-SECTION IN RELATION TO FREQUENCY WITH WHICH DIFFERENT FACES WERE SELECTED BY THE SUBJECT FOR USE AS TARGETS IN TESTS

Target Group	Runs per Trial Position	Hits on 1st Trial	Hits on 6th Trial	Average Diff.	CR_d
Five-Face*.....	104	382	534	1.46	5.77
4-6 Faces.....	102	416	484	.67	2.61
1-2-3 Faces.....	98	390	411	.21	.82

*The five-face was the target for which the subject had stated her preference during the period of the tests.

five-face, which was used most frequently and which the subject said she liked best, showed the most middle-column incline in scoring, with the difference between the first trial position and the last one having a CR of 5.77. The 4-6 faces, intermediate in the

scale of preference as indicated by frequency of use as targets, also showed a significant incline, but the CR in this instance is only 2.61. For the 1-2-3 faces the incline is an insignificant one which has a CR of .82. The new finding of Table 3—one which contributes strongly to the evidence for target preference—is the fact that the five-face sets alone give position effects which are considerably more pronounced than those found in the sets in which the 4-6 faces were the targets. The difference[2] between the two values for the CR_d, 5.77 for the five-face and 2.61 for the 4-6 faces, is such that the likelihood of its being equaled or exceeded by chance is only 14 in 1,000.

The difference between the five-face and the 1-2-3 faces, as shown in Table 3, was evaluated in terms of the observed difference in each set between the number of hits in the combined first trials and those of the combined last trials of the eight middle columns. That is, the difference in question was that between the two groups of sets, 52 for the five-face and 49 for the 1-2-3 faces, as determined from the observed total amount of incline in the eight middle columns of each set. Student's t for the difference between these two distributions was found to be 3.16. Only eight times in 10,000 such sets of data, on the average, would a difference as great as that found here between the five-face and the 1-2-3 faces occur by chance.

Discussion

Before reaching a decision as to whether target preference affected PK performance in these data, we must consider the counter-hypotheses which might be advanced to explain the findings.

The Chance Hypothesis

A brief summary of the results is sufficient to show that the difference in effects associated with the use of the high and the low faces cannot be dismissed as being of a purely random nature. In the U-Section, those sets which were done with the 4-5-6 faces of the dice as targets showed a highly significant incline in scoring within the middle columns (CR = 5.94), while those done with the

[2] I am indebted to Dr. J. A. Greenwood for pointing out that the difference between two CR's of the difference may be evaluated by treating their algebraic difference (that is, keeping the signs for direction of difference) as a normal variate with a variance of 2.

1-2-3 faces as targets showed only an insignificant incline (CR = .82). In the W-Section sets, the 4-5-6 faces showed a significant degree of terminal salience in the middle columns (CR = 2.64), while the other target group showed no such evidence. This latter result was in accordance with a definite prediction based on the findings in the U-Section and made before the analysis of the W-Section was started. It is evident from these critical ratios that significant position effects in the interior of L.H.G.'s sets occurred only with the use of the 4-5-6 faces as targets.

However, when the question was raised as to whether the differences between the two groups of targets were themselves significant, an answer was made in terms of appropriate observed differences in scoring which reduce the scoring trends to quantitative units. This method was applied to each section of the data, U and W in turn. The results of the two independent tests were evaluated by the chi-square method of combination of critical ratios. When the fact was taken into account that the second finding agreed with the first, in accordance with prediction, the difference between the position effects for the two groups of targets was found to be statistically reliable at the one per cent level of significance.

The best case against the chance hypothesis as the explanation of these results is the analysis which at the same time most clearly associates the findings with psychological factors inherent in the subject's reactions to the different targets. In the U-Section data the degree of the psychological effect of the different targets was found to vary directly with the frequency with which the subject elected to use the different faces as targets. The difference between the inclines in scoring for the two extreme groups, that for the five-face on the one hand and that for the 1-2-3 faces combined on the other, gave a Student's t of 3.16 (P = .0008), which effectively rules out the chance hypothesis.

The Biased-Dice Hypothesis

The biased-dice hypothesis in its usual form cannot be applied to explain the differences found between the 1-2-3 and the 4-5-6 faces, for the significance of the findings rests upon hit patterning in the trial positions of the sets when the same targets were used throughout. There is no known way in which physical bias of the

dice could show up in a rhythmic, lawful manner when the same target is used throughout a set. And even if it were possible to suggest some hypothesis based on biased dice to account for the position effects found for the 4-5-6 faces, this would require that the scoring of the 1-2-3 faces show the reverse patterning of hits. Actually, they showed instead a general absence of position effects in the middle columns of the W-Section sets; and in the U-Section they gave an insignificant *incline* in scoring, the same trend as the one which was highly significant for the 4-5-6 faces.

The question of the amount of physical bias of the dice is itself an interesting one, but it is not directly relevant to the conclusion regarding target preference in this paper.

Selection of Data

Can the significant differences found for the target groups be explained on the basis of the selection of the most outstanding results from among many possible comparisons? The answer is that they cannot, for the following reasons. There were, logically, only two parts of the set which were considered to be possible *loci* for the reflection of target preference effects. One was the first column of the set, which showed highly significant declines in scoring in both sections of the work. The other was the middle columns, which gave the characteristic incline in scoring in the U-Section and the curve of both beginning and ending salience in the W-Section. These were the only two places in the data where the effect of target preference was investigated. Furthermore, the expectation from the results of the two M.P.R. series was that there would be little, if any, difference between the high and the low targets in the first column of the set.

The most conclusive argument against the selection hypothesis, however, is the fact that the data as a whole were divided into eight distinct groups before any of the statistical work had been started and the analyses were done independently upon each of these eight batches.[3] The first batch was the only one in which the study

[3] For convenience, the data were first grouped as follows: (A) sets from exclusively U-Section pages; (B) sets from exclusively W-Section pages; (C) U-Section sets from mixed pages; (D) W-Section sets from mixed pages. Then, to find out whether the target effects were consistent at all scoring levels, each group was further subdivided as follows: (1) sets with total scores of 65 and below; (2) sets with total scores of 66 and above. The two score-level groups contributed to an approximately equal degree to the evidence for target preference.

was purely exploratory. On the basis of the outcome of the first batch, it was predicted that the same difference in position effects for the two target groups would show up in the middle columns of the sets, and this was found to be the case, without exception, in the seven remaining batches of the data. This is, in itself, strong evidence of a real difference in effect between the two groups of targets, as the probability that the next seven observations would all agree with the first one is $(\frac{1}{2})^7$ or 1 in 128.

Target Preference or Position of Target on the Page?

In the report on position effects in the last issue, some differences both in general level of scoring and in scoring trends were found

Table 4

FREQUENCY OF USE OF FACES AS TARGETS IN THE DIFFERENT SET POSITIONS ON THE PAGE

Set Number	W-Section				U-Section			
	1-2-3 Faces		4-5-6 Faces		1-2-3 Faces		4-5-6 Faces	
	Times Used	Per cent of Total	Times Used	Per cent of Total	Times Used	Per cent of Total	Times Used	Per cent of Total
1.....	28	27.7	34	20.6	12	24.5	26	25.3
2.....	23	22.8	42	25.5	11	22.4	23	22.3
3.....	20	19.8	36	21.8	7	14.3	21	20.4
4.....	19	18.8	37	22.4	10	20.4	14	13.6
5.....	6	5.9	7	4.2	5	10.2	9	8.7
6-8.....	5	5.0	9	5.5	4	8.2	10	9.7
Total.....	101	100.0	165	100.0	49	100.0	103	100.0

in relation to the location of the sets on the record page. It is important, therefore, to know in connection with the present study how the sets in the two target groups were distributed over the record page. If we find, for example, that the 4-5-6 faces happened to be used as targets more frequently in the more favorable locations on the page, then this fact, rather than target preferences, might be the explanation of the significant differences in position effects. The distribution of the targets on the page is shown in Table 4.

These results show that the 4-5-6 faces were not used any more frequently as targets in the more favorable set positions on the

page. In fact, the situation was more nearly the very reverse of that required by this counterhypothesis. Location of the sets on the page significantly affected scoring only in the W-Section, where the first and third sets on the page, those on the left-hand side, gave more pronounced position effects than did the second and fourth sets, and 47.5% of all the occasions of using the 1-2-3 faces occurred in these sets, as compared with 42.4% for the 4-5-6 faces. Unequal distribution of the targets on the page cannot explain the differences in position effects.

With the counterhypotheses thus found inadequate, we are left with target preference as the only available explanation of the findings.[4]

Was the Target Preference Developed During the Series?

L.H.G.'s experience with dice games had been quite limited before she began serving as subject in the Gibson Cup Series. It seems likely, therefore, that any psychological preference for some of the targets would have been developed sometime during the series. If such distinctive reactions developed very early in the tests, it is entirely possible that their emergence in the results could not now be detected. However, a chronological study of the results for the two target groups was indicated for whatever light it might throw upon this interesting question.

This particular investigation was limited to the 25 pages of the U-Section which contained only unwitnessed sets. The pages were separated into four chronological groups to give the most nearly equal division in terms of number of sets. In these groups, the differences in average between the first and the last trials of the middle columns for the high and the low target groups were observed. The values were recorded as positive numbers when they represented an incline in scoring, the expected direction of difference, and as negative when the combined scoring in the first position of the column was higher than that of the last position. When the 4-5-6 faces were targets, these differences in average for the four quarters of the series were: .90, 1.18, 1.84, and .64. The corre-

[4] A counterhypothesis of variation in the manner of throwing, discussed on pp. 55-63 in this issue, also is found to be inadequate.

sponding figures for the low targets were: 1.81, −.65, .33, and −.50.

It is evident at once that in the first quarter of the series the incline in scoring was just as pronounced with the low targets as with the high ones. Numerically the difference in average was twice as much for the 1-2-3 faces as for the 4-5-6 faces in the early part of the experiment. In the other three-quarters of the series, however, the high targets showed an average incline in scoring of 1.20, while the low targets gave an average decline of .32. The interpretation suggested by these results is that the subject's preference for the 4-5-6 faces was not demonstrated in the first part of the series. In fact, the presence of pronounced position effects from all targets at that stage of the work indicates that all the faces were equally effective as PK targets at first, but that as the series progressed there occurred some kind of inhibition of this psi capacity in the body of the sets when the 1-2-3 faces were used as targets.

Concluding Remarks

The primary purpose of this study has been to determine whether L.H.G.'s relative success in scoring with the different target faces was determined to any degree by anything in the nature of subjective or psychological factors. The existence of distinctive psychological reactions of the subject to the different targets has been established on the basis of consistent and significant position effect differences in relation to the use of the different die faces as the basis of scoring hits in these PK tests. Following a precedent which has been set by earlier PK reports, this psychological cause for differentiation among the different faces has been referred to throughout this paper as "target preference." It should be understood, however, that this term is used more as a convenience than as an explanation. There is no way of knowing at present just what causes the subject's different reactions to the different faces. Perhaps considerable further work will be necessary to find out, and in all likelihood further research along the lines of specifically designed experiments will be required. For the present, the safest course is to recognize that when the psychological effect which has been discovered in these data is attributed to target preference, we are dealing with it at the descriptive level only.

Perhaps this point can be made most effectively by asking a few of the many questions which remain to be answered. What percentage of PK subjects show a differentiation among their targets on a psychological basis? To what extent can target preference in different subjects be explained in general terms, and to what extent is it an individual matter? Is the effect indeed one of inhibition of scoring on the less favored targets rather than of facilitation on the "preferred" ones, as the findings of the previous section suggest? Is the effect of preference on PK results one of which the subject is conscious at the time it occurs? Can the effect be changed by volitional effort on the part of the subject? Can it be influenced by either waking or hypnotic suggestion? Only as these questions and others like them are answered through further research will we reach a more complete understanding of target preference and of its role in PK tests.

References

1. Dale, L. A. The psychokinetic effect: the first A.S.P.R. experiment. *J. Amer. Soc. psych. Res.,* 1946, **40,** 123-51.
2. Gibson, E. P., Gibson, L. H., and Rhine, J. B. A large series of PK tests. *J. Parapsychol.,* 1943, **7,** 228-37.
3. ———. The PK effect: mechanical throwing of three dice. *J. Parapsychol.,* 1944, **8,** 95-109.
4. Pratt, J. G. Lawfulness of position effects in the Gibson Cup PK Series. *J. Parapsychol.,* 1946, **10,** 243-68.
5. Reeves, M. P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.,* 1943, **7,** 76-93.
6. ———. The PK effect: the first doubles experiment. *J. Parapsychol.,* 1945, **9,** 42-51.
7. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.,* 1943, **7,** 20-43.

Parapsychology Laboratory
Duke University
Durham, North Carolina

MINOR ARTICLES

A PK EXPERIMENT UNDER LIGHT AND DARK CONDITIONS

By ELIZABETH A. MCMAHAN

ABSTRACT: In a disc experiment in PK, children and adolescents were tested in a social atmosphere under the conditions of light and dark. A positive deviation was obtained in the dark and a negative one in the light. The difference in scoring level under the two conditions was great enough to give a suggestive, though not quite significant, probability of .014. A probable explanation for the scoring trends is given in terms of the novelty effect of darkness.—Ed.

FROM TIME to time in parapsychology literature suggestions have been made concerning probable effects of lighting conditions on psi processes. In ESP literature (8) there have been references to subjects' endeavoring to shut out sensory stimuli by closing their eyes, staring at a blank surface, or looking off into the distance. These adaptations probably represent an effort to eliminate the distraction caused by visual stimulation. At the seances of Spiritualist mediums it is more or less accepted practice to require darkness or dim red illumination. One of the more recent investigations dealing with the effect of light is a PK experiment reported by Gibson and Rhine (2). In this experiment the subject made a significantly negative score when throwing dice in the dark and scored only at a chance level in the light. It was therefore considered worth while to introduce the variable of differential lighting into a disc-throwing experiment in PK.

The experiment reported here is the fourth in a series of investigations in which discs were used to test the PK hypothesis. The preceding experiments have already been published (4, 5).

PROCEDURE

The present work was carried out between April 28 and September 3 of 1945. Like two of the previous experiments (II and III), it was conducted as a series of "PK parties," and the subjects,

who were adolescents and children, were tested in a social situation. All the subjects were girls. The only innovation consisted in alternating light and dark conditions during the testing.

The five sessions composing the present experiment (Experiment IV) were held at night in order more easily to control the lighting. The sessions usually began at eight o'clock, and each lasted approximately one and one-half hours. Four or six individuals were invited to each party so that there could be an equal number of runs for both targets under the two variables, light and dark.

The apparatus and the record sheets were similar to those used previously. The discs were of a white plastic material, ⅞ of an inch in diameter and 1/10 of an inch thick. On one side of each disc a blue cross was painted, and on the other, a blue circle. Ten discs were thrown at a time by means of a rotating wooden tube, 5½ feet long, which had a flat-bottomed, wire, basket-like structure on either end to catch the discs. This disc cage has been described and pictured in a previous report (**4**).

The record sheets had four columns across the page with spaces for five entries in each column. These four columns composed a set. In a single turn a subject completed one of these sets, rotating the cage while the other subjects gathered about the apparatus and "helped" her concentrate on the target. Rewards of candy, toys, and movie tickets were given for good scores, and refreshments were served at the end of each session.

A die throw determined whether the first subject's turn would be taken in the light or in the dark. Thereafter the conditions were alternated until every subject had had one turn. Then, in the same order, each subject was tested under the condition opposite to that of her first turn.

Die throws also determined the order in which the subjects took their turns, as well as the target face for each subject, except when equalization of the targets (for control on biased discs) necessitated the choosing of a certain face.

Under conditions of "light" an ordinary overhead light was on. Under "dark" conditions the light was out and the room was so darkened that the subjects could not see one another.

In throwing the discs under "light" conditions a subject would rotate the cage until it was in a vertical position and the basket containing the discs was at the top. The discs tumbled down over the baffles which lined the chute and came to a stop on the bottom of the lower basket. Here they were counted, and both the number of hits and the number of misses were called aloud and recorded by the experimenter.

For a "dark" turn the lights were extinguished, and the subject then rotated the cage. When all the discs had come to rest in the lower basket, a flashlight was held on them until the hits and misses had been called and recorded. The flashlight was then switched off and the cage was given another half-rotation.

Dr. Betty M. Humphrey cooperated with me in this as in all the other group disc experiments.

The limit to the experiment was set, after the end of the second session, at a total of five sessions.

Results

The runs[1] and deviations for the two conditions of "light" and "dark" are represented in Table 1 for the five sessions.

Table 1

General Results

Session	Number of Subjects	Light		Dark		CR_d
		Runs	Dev.	Runs	Dev.	
1 (Children)....	4	40	−12	40	+38	2.50
2 (Adolescents).	4	40	−11	40	+17	1.40
3 (Children)....	4	40	−22	40	− 4	.90
4 (Children)....	4	40	+ 6	40	+ 3	.15
5 (Adolescents).	6	60	−22	60	0	.90
Total.....	22	220	−61	220	+54	2.45 2P = .014

When the data for all the subjects are pooled, it will be noted that the subjects scored positively under dark conditions and negatively in the light. The CR of the difference between the two con-

[1] The standard run in a disc experiment is 20 discs. In the present experiment two throws of ten discs composed a run, and each subject, by making 20 throws in a turn, completed 10 runs per turn.

ditions is 2.45 which is significant, having a probability of .007. This should be corrected to .014 to allow for direction of deviation.

As one can tell from the figures in Table 1, the total deviation for the entire experiment is not significant (−7 on 440 runs).

In the previous disc experiments there was a rather striking and significant decline between the first and second turns when the data for all subjects were pooled. This decline is found in the present work as well, although it is not significant. The data are given in Table 2.

Table 2

DECLINE IN SCORING BY TURNS FOR THE SESSIONS OF EXPERIMENT IV

Session	First Turn		Second Turn		CR_d
	Runs	Dev.	Runs	Dev.	
1	40	+21	40	+ 5	.80
2	40	+ 7	40	− 1	.40
3	40	+ 4	40	−30	1.70
4	40	−10	40	+19	1.45
5	60	− 1	60	−21	.82
Total	220	+21	220	−28	1.04

By pooling the data for disc Experiments II and III with those of Experiment IV (Experiment I is omitted because each subject had only one turn) a CR of the difference between turns one and two amounting to 3.21 is obtained. This has a probability of .0007, which is highly significant. When the CR's of the difference for the individual experiments are combined by the chi-square method of evaluation, the resulting P is .005. Table 3 shows the results of the two-turn analysis for the individual experiments and for all of them together.

Table 3

DECLINE IN SCORING BY TURNS FOR EXPERIMENTS II, III, AND IV

Experiment	First Turn		Second Turn		CR_d	CR^2
	Runs	Dev.	Runs	Dev.		
II	337.5	+113	337.5	−88	3.45	11.90
III	160	+ 45	160	+20	.62	.38
IV	220	+ 21	220	−28	1.04	1.08
Total	717.5	+179	717.5	−96	3.21 P = .0007	χ^2 = 13.36 P = .005

When the data for Experiment IV were analyzed for position effects on the record sheet, no significant trends were discovered. The horizontal distribution of hits for the four columns across the page was as follows, with each of the upper four figures representing 110 runs:

$$\underbrace{-10 \quad -12}_{-32} \qquad \underbrace{+9 \quad +16}_{+25}$$

The vertical distribution for the five entries in the columns showed the following deviations, each deviation representing 88 runs:

$$\left.\begin{matrix}+10\\+14\end{matrix}\right\} +24$$
$$-26$$
$$\left.\begin{matrix}+18\\-23\end{matrix}\right\} -5$$

Neither of these distributions shows a significant difference, and it is evident that a quarter distribution analysis would not be expected to yield a typical QD.

The QD's of the set are given below for the "light" and "dark" conditions, both separately and in combination. None of the QD's are significant.

Light

44	
-3	0
-16	-1

Dark

44	
+2	+25
-1	+13

Total

88	
-1	+25
-17	+12

For the sake of completeness the QD of the set is given below for all the disc work I have completed up to the present. It is not significant although fairly close to the typical QD's found in dice work.

Total QD for Experiments I, II, III, IV

694.5 +108	+20
-42	-23

$CR_{d(1-4)} = 1.57$

Discussion

Before the results of the present experiment can be attributed to PK, the counterhypotheses must be eliminated as possible explanations. Those of biased discs, skilled throwing, and optional stopping were adequately met in the experimental set-up. The data have been rechecked and recomputed, and no errors have been found. The CR of the difference between the light and dark conditions (after due corrections are made) gives a probability of .014. This does not eliminate the chance hypothesis, but makes it rather unlikely. PK appears to be the best explanation.

The results of this experiment show that the scoring was facilitated by the condition of darkness. The best explanation of such a result would seem to lie in the motivation of the subjects themselves, who, as a whole, preferred the dark. Both Dr. Humphrey and I noticed this fact. In the first session, even before an appreciable difference in scoring rate was observed, the subjects' comments indicated such a preference. The following comments are taken from notes made at the first session: "After the first few turns, the subjects made remarks indicating their preference for the dark condition. Those who had had 'light' for their first turn were glad their second turns would be 'dark.' They got the idea that 'dark' was lucky." This attitude of liking for the dark condition was noticed in the other sessions also.

Perhaps the reason for such an attitude was the fact that being tested in the dark constituted a new situation for the subjects. The favorable effect of novelty upon scoring, both in ESP and PK tests, has been often commented upon in psi literature (**1**; **3**; **7**; **8**, p. 225; **9**, pp. 281-82; **10**).

It was something of an adventure to sit in total darkness and listen to the discs come rattling down the chute. Being tested under light conditions, on the other hand, was an old thing to most of the subjects. They had been present at several previous PK parties at which all the testing had been done in the light. Even though these same children had enjoyed the tests they had taken under light conditions on previous occasions, they now looked upon light as "unlucky" and compared it unfavorably with the dark condition. Pegram (**6**) has noticed that when two conditions, both of which are favorable but unequally so, are alternated, the results indicate that the less preferred condition is disfavored by comparison, which results in a lowering of the scoring rate. Perhaps this explains why the deviation for "light" is −61 rather than a figure nearer the chance level.

The conflict between the Gibson findings (in which the subject made a significantly negative score with dice in the dark and scored near chance both in red light and in white) and those of the present report makes it necessary to look for reasons for the discrepancy. There are at least two factors which were quite different in the two experiments. In the first place, the Gibson subject (Mrs. Gibson) was an adult, while the subjects in the disc experiment ranged in age from eight to seventeen years. It is quite conceivable that to an adult, the process of throwing dice in the dark would not long be exciting but would soon become tedious because of the necessary slowing-up of the testing procedure. (It is interesting to note that in the Gibson data the first three columns of data done in the dark have a higher average deviation than have the first three columns done in either white light or red light. This initial advantage, however, is lost rather rapidly over the remaining nine columns.) Mrs. Gibson was tested alone, while the children were tested in a social atmosphere.

At this point it may be of interest to consider one of the sessions in a little more detail because of certain factors which were peculiar to it. Table 1 shows that the only session in which the average scoring rate for the light conditions was higher than that for the dark conditions was Session 4. The subjects in this session were children ranging in age from eight to thirteen years. The

youngest subject was eight-year-old Carolyn who was nearly three years younger than the next oldest. Unlike any of the others she had never been to a "PK party" before and did not know just what to expect. She won the right to the first turn of the session by throwing a higher number on a die than did the other subjects. Another die throw determined that the beginning turn should be under dark conditions. Perhaps these facts offer a possible explanation as to why Carolyn's first turn ("dark") gave a negative deviation of 8 while her second turn ("light"), taken after half the session was over and when she was more accustomed to the procedure, gave a positive deviation of 14. The other three subjects in the session scored better in the dark.

If Carolyn's data are omitted from the results, the light-dark division for session 4 is as follows:

Light		Dark	
Runs	Dev.	Runs	Dev.
30	−8	30	+11

The omission of Carolyn's data from the total results of all five sessions pooled gives these figures:

Light		Dark	
Runs	Dev.	Runs	Dev.
210	−75	210	+62

$$CR_d = 2.98$$
$$P = .0015$$

These observations are given for what they are worth; it is not intended that they should alter the general conclusions reached in this paper. They are of interest, however, in the interpretation of the experimental results, especially with regard to future experimentation. If we could be sure that Carolyn's inexperience was the true explanation of her divergent results we could safely attribute importance to the smaller probability of .0015. This explanation, however, is only a probable and not a necessary one.

It should be recognized that there is another possible explanation for the fact that dark conditions were better than light in the present experiment. It is that darkness shuts out many distractions and enables the subject more easily and effectively to concentrate his attention upon the test. This explanation, however, does not appear to fit the facts as well as does the one of novelty. First of

all, the "dark" data are roughly comparable, in terms of average deviation, with the data obtained under the best "light" conditions of previous experiments. In the present "light" tests, the data are negative and as such are on a scoring level with results obtained under the lowest scoring with light in earlier experiments. The relationship between degree of lighting and scoring level, then, is a relative one.

It must also be kept in mind that any hypothesis stating that darkness favors scoring conflicts with the Gibson data.

The present experiment supports the evidence that the conditions affecting the interest and attention of the subject are extremely important. The same physical conditions in two test situations may produce different or even opposite results if the motivation of the subjects in the two situations is different.

References

1. Estabrooks, G. H. A contribution to experimental telepathy. *Bull. Boston S.P.R.*, 1927, **5**, 1-30.
2. Gibson, E. P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, **7**, 118-34.
3. Guthrie, M. Further report on experiments in thought transference at Liverpool. *Proc. S.P.R.*, 1885, **3**, 424-52.
4. McMahan, E. A PK experiment with discs. *J. Parapsychol.*, 1946, **10**, 169-80.
5. ———. PK experiments with two-sided objects. *J. Parapsychol.*, 1945, **9**, 249-63.
6. Pegram, M. A comparison of ESP test procedures. [Unpublished MS.]
7. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extrasensory perception. *J. Parapsychol.*, 1939, **3**, 121-58.
8. Rhine, J. B. *Extrasensory Perception.* Boston: Bruce Humphries, 1934.
9. Rhine, J. B., Pratt, J. G., Stuart, C. E., and Smith, B. M., with Greenwood, J. A. *Extrasensory Perception after Sixty Years.* New York: Henry Holt, 1940.
10. Rhine, L. E. Some stimulus variations in extrasensory perception with child subjects. *J. Parapsychol.*, 1937, **1**, 102-13.

Parapsychology Laboratory
Duke University
Durham, North Carolina

DICE BIAS AND MANNER OF THROWING

By J. G. Pratt

ABSTRACT. This paper reviews the methods which PK experimenters have used to eliminate the hypothesis of imperfect dice as the explanation of significant results in their tests. It also raises the question whether the bias of the dice might influence the results to different degrees in relation to the amount of effort, or to the rhythm used, in throwing them. An extensive control series was conducted in which three contrasting ways of throwing dice with excavated spots were compared. The dice were found to be biased to favor the occurrence of the three higher faces, regardless of the amount of effort used in throwing them. No significant differences among the three styles of throwing were found. The conclusion indicated is that position effects in PK test results cannot be explained by a combination of biased dice and systematic changes in the manner of throwing the dice.—Ed.

Introduction

The conclusion that target preference is a factor which sometimes affects PK control over the fall of dice was reached in a report appearing elsewhere in this issue. That conclusion did not depend in any essential respect upon the question of the degree to which the dice actually were biased. It is important, however, to clarify the situation with respect to these two possible variables as fully as we can.

From the point of view of the PK research worker, physically "true" dice do not exist. The tests have always been planned to provide their own controls against imperfect dice, and the degree of bias of the dice was not a question of great importance. All that mattered was that the cubes be sufficiently true to give results which could be evaluated statistically. Ordinary commercial dice with excavated spots met this requirement, and they were used in many tests as a matter of convenience. Theoretically, however, dice with excavated spots should, from physical considerations alone, favor the higher faces. Many of the PK reports have presented results which fit into this pattern. In tests based upon combinations of faces, high dice have averaged above low dice (**7, 12**), and double

sixes have been better than double ones (**8**). Also, of those singles tests which followed the procedure of throwing "around the die," the majority show a trend for the rate of success on the different targets to be directly related to the numerical value of the faces (**1, 2, 3**).

The PK investigators, in planning their experiments to eliminate the biased-dice hypothesis, used three methods of control by design, and a fourth emerged during the analysis of the results after most of the data had been collected. (1) The same dice have been used under identical test conditions with high and low faces or combinations as targets, so that the physical imperfections which favor some faces would be cancelled out by disfavoring others. When the results of all faces combined are significantly higher than the expected chance average, the hypothesis of biased dice is effectively eliminated. (2) Some tests have been conducted with only the six-face as target because of the subjects' expressed preference for this face, but these have usually involved throwing under two or more different psychological conditions. The test of significance in these instances has rested upon the evaluation of differences in rate of scoring upon the six-face in different parts of the experiment. Where such comparisons show that significant changes in rate of scoring were obtained with the same dice thrown for the same face, the biased-dice explanation cannot be applied. (3) A few investigators have run "control" series with the dice to be used in PK tests. In these series they have recorded the number of times the various faces appeared uppermost when no conscious effort was made to influence them mentally. Some of these control series have helped the investigators to conclude that the dice were adequately true for their purposes (**4, 5, 9**). In such a control series there is always present, however, the theoretical possibility that the thrower is influencing the dice by PK whether he is consciously setting himself to do so or not. This causes no difficulties as long as the research is concerned with the problem of evidence that PK occurs, for all that matters is whether the difference is significant. (4) The best answer to the biased-dice hypothesis came out of the position effect analyses. The bearing of the results of the QD analyses, for example, on the question of biased dice is direct and decisive.

To invoke the physical explanation to cover the variations in hits occurring with the same target in different parts of the record structure would be to suppose that the effect of the bias of the dice followed a rhythmic pattern. However, even this remote logical possibility was considered by the investigators, since they felt obliged to explore every conceivable counterhypothesis.

In two reports dealing with PK tests, with the six-face alone as the target, in which both significant total deviations and significant QD declines were found, Rhine and Humphrey discussed what they called the "combination hypothesis" as an alternative explanation to PK (**10, 11**). This suggestion makes use of a combination of physical assumptions: (1) that the dice are biased to favor the higher faces, the six-face most of all, because of the lightening effect of the cavities; (2) that this bias will show up in proportion to the distance the dice are thrown on a roll; and (3) that the subjects characteristically put more force into their throwing of the dice at the start of a new page or set, thus getting a higher deviation on the target face in the first quarter than in the fourth one.

There is a good theoretical basis for the first of these assumptions (that is, that excavated dice are biased); but the second one is certainly debatable, and there is no evidence for the third one. For the combination hypothesis to be valid, all three conditions would have to be established. In any case, the combination hypothesis, even if it were substantiated, could not explain the QD results as a whole. Rhine and Humphrey point out that in many series with significant QD results the dice were tumbled in motor-driven cages or were released mechanically to fall under the force of gravity. These experimental conditions eliminate the possibility of rhythmic changes in the manner of throwing. They therefore rule out the combination hypothesis by violating one of its basic assumptions.

In a discussion of the results of the target preference study of L.H.G.'s results, it was suggested that the combination hypothesis in a modified or more extreme form might explain the difference in position effects between the two groups of targets. The hypothesis was not offered in this case as an alternative to PK, for

this is sufficiently established by the position effects for all faces combined, particularly by the incline in success in the middle columns in the unwitnessed (U-section) sets. But the rhythmic-throwing-plus-biased-dice hypothesis was proposed as offering a possible alternative to target preference as an explanation of the difference in position effects among the groups of target faces investigated. Unlike the previous reports wherein the combination hypothesis was considered as a counterhypothesis to PK itself, this possible effect of biased dice was offered here only as an alternative to target preference. The evidence for PK in this series was not questioned.

The suggestion made took the following form: The manner of throwing at the start of the middle columns in the U-Section was not adequate for the bias of the dice to assert itself in any way; also, PK was inhibited by the conditions prevailing for those particular throws. Therefore all target faces scored at "chance." As the subject proceeded through each column, the general level of scoring was raised by the fact that the conditions became more favorable for PK; at the same time, a change in the rhythm or manner of throwing permitted the loading of the dice, which favored the higher faces, to show up.

This hypothesis becomes so complicated and it has so many facets that it seems to the writer far less reasonable on purely logical grounds than the explanation that the subject's PK ability was affected by different psychological reactions to the targets. However, it did suggest the possibility of gathering some empirical data to bear upon the question of whether success on the different faces actually varies under sharply contrasting methods of throwing. In any such control series there is, of course, always the question of whether the PK factor has been fully excluded. It is recognized that the steps taken in this investigation may not have ruled out the *possibility* of PK, though they appear adequate in the light of present knowledge.

Procedure

Six common white dice having excavated spots marked with black paint were used. The dice originally thrown in the Gibson Cup Series were not available, but the ones selected were considered

adequate to show any systematic bias attributable to the effect of the cavities on the faces. These dice were thrown, six at a time, from an ordinary drinking glass with smooth, tapered sides, onto an experimental dice table with padded and felt-lined top and sides. The glass measured approximately two inches in diameter at the bottom, three at the top, and five inches in depth. The dice were $\frac{3}{16}$ of an inch on each edge.

Three types of throwing were used, representing the possible variation in rhythms which a subject might use in any test situation in which the dice were thrown by hand and recovered by the subject without any necessity for moving from one position.

The "No Roll" condition consisted of picking the dice up from the table as they lay from the preceding throw and placing them in the glass as easily as could be done with convenience, then simply tilting the glass up with the lip resting on the surface of the table until the dice slid out by gravity. In this condition, the dice were given only the minimum amount of movement necessary to transfer them from the table to the glass and back again to the table for scoring purposes.

In the "Long Roll" condition, the dice were handled with vigor as they were picked up and dropped into the glass, and they were thrown as far as possible on the dice table without hitting the wall except in rare instances, an average roll of about three feet on the padded surface.

In the "Hard Baffle Bounce" condition, the greatest amount of effort was put into the throwing, and the dice were always bounced with force off the end wall of the table.

In order to have an observer who was unaware of the purpose of the control series so as to minimize the possibility that he would influence the results by PK, the services of Mr. J. A. Hornaday, Jr., graduate student in the Department of Psychology at Duke University, were obtained to collect all the data. He had not been present at any of the discussions of the possible effects of biased dice in relation to methods of throwing which took place among members of the Parapsychology Laboratory in connection with the target preference findings reported in this issue. When he was questioned at the end of the series, he had not surmised correctly the bearing of the data.

One thousand throws were made under each condition. The length was fixed before the start of the control series. The distribution of faces for all six dice was recorded at each throw. The observer worked under pressure to complete the task, which required 15 hours of working time, interrupted only for eating and sleeping. These conditions would be generally accepted by research workers as unfavorable to the occurrence of psi phenomena, and they render it quite unlikely that PK was a factor in producing the results obtained. That a high standard of accuracy was maintained throughout in making the observations is shown by the fact that the correct total of 6,000 die readings was recorded for each condition. The only errors possible, therefore, are those of misreading faces, and there is no reason to believe that these instances, if they occurred, would have been of other than random nature, so they would have cancelled each other out.

Results

The results of the control series are given in Table 1. In data such as these, in which all of the dice are recorded on each throw, the 1-2-3 faces and the 4-5-6 faces necessarily deviate from "chance" by the same amount but in opposite directions. In order to have statistically independent data for evaluation, we may base the comparison of the different ways of throwing the dice upon the frequencies for the 4-5-6 faces alone.

Inspection of the figures in the table shows that these results followed the assumption generally made that excavated dice will give a greater frequency of the higher faces.

Table 1

Effect of Manner of Throwing on Face Distribution

Throwing Condition	Deviations from "Chance"*						4-5-6 Faces Pooled		CR of Largest Difference (No Roll vs. Hard Bounce)
	One-Face	Two-Face	Three-Face	Four-Face	Five-Face	Six-Face	Dev.	CR†	
No Roll.....	−76	−59	− 2	24	62	51	137	3.54	
Long Roll....	−93	−37	29	7	57	37	101	2.61	1.30
Hard Baffle Bounce....	−67	−26	27	34	−18	50	66	1.70	

*A total of 1,000 throws, or 6,000 single die falls, were observed under each condition, giving 1,000 as the average chance expectation for each face if the dice are true and if random factors alone are involved.

†CR = Dev. / $\sqrt{n\,p\,q}$ = Dev. / $\sqrt{6{,}000 \cdot \frac{1}{6} \cdot \frac{5}{6}}$

The "No Roll" condition showed the greatest deviation of the 4-5-6 faces, which totaled 137 more than chance expectation. The "Long Roll" was next, with a positive deviation of 101 on these faces, and the smallest number of 4-5-6 faces, 66 more than the expected chance average, was observed on the "Hard Baffle Bounce" condition. *None of the differences among the deviations for these three groups are significant, however, so there is no ground for concluding that they represent more than random variation in the fall of the dice as far as any effects traceable to the three methods of throwing are concerned.* This conclusion is strengthened by the fact that the biggest deviation on the 4-5-6 faces occurred with the "No Roll" condition, which is where it would be least expected if the distance or force of throwing had any appreciable effect.

Discussion

Perhaps there is some need for caution in interpreting the results of this series, just as there has been for other controls with dice in which the object was to eliminate the PK factor. The difficulty is in knowing with certainty just what psychological conditions would effectively free the results from any influence of this psi function. The discovery of evidence of PK in such subtle forms as position effects which occur without the awareness of the subject and which are influenced by such factors as the structure of the records, the position of sets on the page, and target preference suggests that controls against PK may fail in ways which still elude discovery.

It is not possible, therefore, to conclude with certainty that the thrower did not use PK to influence the fall of the dice in this control series. At best, we can only state again that the conditions seemed to come as close to providing a test of the purely physical influences on the action of the cubes as it is possible to get as long as there is any element of human motivation in the situation at all. If ever it is desirable to obtain more definite controls, it would appear necessary to introduce some method of throwing and recording the dice by mechanical and completely automatic means. By working out the necessary refinements of procedure along these

lines, it should be possible to reduce the likelihood of any PK influence in the results to such small proportions that it could safely be ignored.

Nevertheless, it is safe to say that it is highly probable that the results of the control series reported here are attributable to the physical factors in the situation, and not to PK. The conclusions to which this evidence points are as follows: (1) Ordinary dice with excavated spots are physically biased to such a degree that their manner of coming to rest after being thrown favors the higher faces to a measurable degree. (2) Within the limits tested, broad enough to include most PK tests in which the dice are thrown by hand, the manner or rhythm of throwing does not affect the degree to which the bias of the dice shows up in the distribution of faces.

The control series therefore suggests that, in the many experimental series with excavated dice which have given higher scores on the higher faces, the bias of the dice was a contributing factor. The findings of the target preference study suggest that the subjects' psychological reactions to the faces of the dice may at the same time have influenced the scores in some of these series.

References

1. Dale, L. A. The psychokinetic effect: the first A.S.P.R. experiment. *J. Amer. Soc. psych. Res.*, 1946, **40**, 123-51.
2. Gibson, E. P., Gibson, L. H., and Rhine, J. B. A large series of PK tests. *J. Parapsychol.*, 1943, **7**, 228-37.
3. ———. The PK effect: mechanical throwing of three dice. *J. Parapsychol.*, 1944, **8**, 95-109.
4. Hilton, H., Jr., Baer, G., and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 172-90.
5. Humphrey, B. M., and Rhine, J. B. PK tests with two sizes of dice mechanically thrown. *J. Parapsychol.*, 1945, **9**, 124-32.
6. Pratt, J. G. Lawfulness of position effects in the Gibson Cup PK Series. *J. Parapsychol.*, 1946, **10**, 243-68.
7. Reeves, M. P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7**, 76-93.
8. ———. The PK effect: the first doubles experiment. *J. Parapsychol.*, 1945, **9**, 42-51.

9. RHINE, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7**, 207-17.

10. RHINE, J. B., and HUMPHREY, B. M. Position effects in the Six-by-Six Series of PK tests. *J. Parapsychol.*, 1945, **9**, 296-302.

11. ———. The PK effect with sixty dice per throw. *J. Parapsychol.*, 1945, **9**, 203-18.

12. RHINE, L. E., and RHINE, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

Parapsychology Laboratory
Duke University
Durham, North Carolina

LETTERS AND COMMENTS

THE PHYSICAL SIGNIFICANCE OF PSI EXPERIMENTS

From the experiments in ESP and the subsequent research on precognition and the psychokinetic effect we have proof of the existence of a quite unforeseen aspect of nature. Although it is the psychologists who are most interested in psi phenomena and who have played a dominant part in the research work, we must not forget that we are dealing also with physical relations. In ESP and precognition a gifted subject works, on the whole, as a registrating apparatus for physical facts. In the psychokinetic effects (I assume their genuineness) again we encounter an entirely new kind of physical reaction.

That the way from ESP to precognition and psychokinesis was quite logical is evidenced by the fact that simultaneous and independent experiments on these matters were undertaken all over the world, once the classical books on ESP were published. For the better understanding of the facts the experimenters tried to get different effects by varying the factors of space and time, but, as the results stand today, without any remarkable success. It follows that we cannot assume in these experiments that energy is transmitted in the ordinary way known to normal physics, because the latter nearly always reveal a decrease of intensity proportional to the reciprocal of the square of the spatial distance. Exact proof of the freedom from space and time we can only get from experiments in which knowledge of the position of the target object or telepathic agent is withheld from the subject, so as to exclude eventual subconscious directing of the transmitted energies. Such questions are but of secondary importance, however, since we may now take precognition for granted. Precognition introduces a deciding fact into any physical mode of viewing things. Physics must in this matter clear up two questions: first, whether these effects will be taken under consideration at all; and second, whether the known laws of physics admit even the possibility of such effects.

If we consider the task of physics to be restricted to research into laws which causally or statistically determine the behavior of matter, there will, of course, be no need for it to consider parapsychic phenomena as such, and we can leave such experiments to a science which is especially to be built up for the purpose. But as space, time, and matter are involved in a new way in these experiments, and as we have great difficulty in physics to understand these concepts and conduct fundamental experiments with them, it will hardly be possible for any physicist to keep away from parapsychological findings. It is not impossible that they will provide ideas for future physical research on a broader front.

In ESP we do not at first find essential difficulties because the facts can, of course, be stated as established even though we may not understand how they come about. In precognition, facts are stated which are not facts from the point of view of statistical physics, which negates on principle any possibility of predicting on the basis of common statistical probabilities when or if there is no assignable causal determination. In psychokinetic effects we do not know how energy is transmitted or whether energy is transmitted at all in the common sense.

If we try to approach these psi effects from the physical side, we find the only way to them is where causal determination ceases, that is, in microphysics. Heisenberg's law of uncertainty between conjugated variables, such as position and momentum or time and energy, leaves open to chance under special conditions the direction of the movement of a small particle of matter. This view has already been used by Bohr, Jordan, and others for an approach to biology from the physical side.

Just as in microphysical happenings there are left open to chance quite a lot of possibilities, we can assume in PK a switching apparatus working without energy and causing PK effects. We see a special importance in energyless functioning because any use of energy involves time as a conjugated variable, and in precognition we have to get out of time in any case. As a logical consequence we must assume some nonclassical stratum which connects past, present, and future and is accessible to the conscious mind in a passive way (ESP, precognition) and in an active way (PK).

There seems to be no need for an uncompromising dualism in order to attempt the study of the properties of the above-mentioned stratum by experiments from the physical and perhaps from the biological side.

Physics will probably keep out of the main line of these investigations, but will surely not be left uninfluenced by the concrete results of the new science.

F. I. Havlicek
Zagreb, Jugoslavia

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this JOURNAL, the following definitions are submitted for convenient reference.

*For a simple description of those terms marked by an asterisk, as they apply to the ESP test data, see Chapter VIII and the Appendix of *A Handbook for Testing Extra-Sensory Perception* by C. E. Stuart and J. G. Pratt. Further explanation may be found in any elementary statistical text.

AGENT: In tests for telepathy, the person whose mental states are to be apprehended by the percipient. In GESP tests, the person who looks at the stimulus object.

AVERAGE SCORE: Average number of hits per run.

CALL v.: To attempt to identify by ESP a stimulus object or a mental state of an agent.

CALL n.: The response described above; also the resulting selection.

*CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.
CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely score if only chance obtains.
CHANCE AVERAGE: Mean chance expectation expressed as an expected score, generally in terms of average per run.

CHI-SQUARE: A sum of quantities each of which is a deviation squared divided by an expected value. Also a sum of the squares of CR's.

CLAIRVOYANCE: Extrasensory perception of objective events as distinguished from telepathic perception of the mental state of another person.

*CR (CRITICAL RATIO): A measure to determine whether or not the observed deviation is significantly greater than the expected random fluctuation about the average. The CR is obtained by dividing the observed deviation by the standard deviation. (The probability of a given CR may be obtained by consulting tables of the probability integral, such as Pearson's.)
CR OF THE DIFFERENCE: The observed difference between the average scores of two samples of data divided by the standard deviation of the difference.

DECK: Twenty-five ESP cards, five of each suit.

*DEVIATION: The amount an observed number of hits or an average score varies from the mean chance expectation or chance average. A deviation may be total (for a series of runs) or average (per run).

DIE THROW: The elementary unit of data in a dice-throwing test represented by the throwing and reading of a single die, no matter how many dice are thrown at a time.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of the hypothesis being tested.

ESP (EXTRASENSORY PERCEPTION): Response to an external event not presented to any known sense.

ESP CARDS: Cards, each bearing one of the following five symbols: star, circle, three parallel wavy lines (called "waves"), square, plus.

ESP SYMBOLS: See plate opposite page 1, this JOURNAL, Vol. 1, March, 1937.

EXPECTATION; see CHANCE.

EXTRACHANCE: Not due to chance alone.

FREE MATERIAL: The stimulus objects of experiments in which an unlimited or unspecified range of stimulus objects is employed (as contrasted with methods such as card-calling in which the subject knows that the stimulus object is one of a known range).

FREE RESPONSE METHOD: An ESP test method in which a relatively unlimited range of stimulus objects is used and in which the subject is instructed to express freely as his response any range of thoughts and ideas.

GESP (GENERAL EXTRASENSORY PERCEPTION): A technique designed to test the occurrence of extrasensory perception, permitting either telepathy or clairvoyance or both to operate.

HIGH-DICE TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling eight or more.

LOW-DICE TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling six or less.

MEAN CHANCE EXPECTATION; see CHANCE.

*P (PROBABILITY): A mathematical estimate of the expected relative frequency of a given event if chance alone were operative.

PARAPSYCHOLOGY: A division of psychology dealing with those psychical effects which appear not to fall within the scope of what is at present recognized law.

PERCIPIENT: The person who makes the calls in an ESP test.

PK (PSYCHOKINESIS): The direct influence exerted on a physical system by a subject without any known intermediate physical energy or instrumentation.

PRECOGNITION: Cognition of a future event which could not be known through rational inference.

PREFERENTIAL MATCHING: A method of scoring free responses. A judge ranks the stimulus objects (usually in sets of four) with respect to their similarity to, or association with, each response; and/or he ranks the responses with respect to their similarity to, or association with, each stimulus object.

PSI: A general term to identify personal factors or processes in nature which transcend accepted laws. It approximates the popular use of the word "psychic" and the technical one, "parapsychical."

PSI PHENOMENA: Occurrences which result from the operation of psi. They include the phenomena of both ESP (including precognition) and PK.

QD (QUARTER DISTRIBUTION): The distribution of hits in the record page (or in a logical subdivision thereof, such as the set or the half-set) as found in the four equal quarters formed by dividing the selected unit horizontally and vertically.

RUN: A group of trials, usually the successive calling of a deck of 25 ESP cards or symbols. In PK tests, 24 single die throws regardless of the number of dice thrown at the same time.

SCORE: The number of hits made in one run.

TOTAL SCORE: Total of scores made in a given number of runs.

AVERAGE SCORE: Total score divided by number of runs.

•SD (STANDARD DEVIATION): Usually the theoretical root mean square of the deviations. It is obtained from the formula $\sqrt{npq}$, in which n is the number of single trials, p the probability of success per trial, and q the probability of failure. (For ESP cards, SD $= 2\sqrt{\text{no. of runs.}}$) Sometimes the SD used is that estimated from the observed variability in the scores.

SD OF THE DIFFERENCE: For both ESP cards and PK tests using dice, the SD of the difference is equal to $\sigma_s \sqrt{1/R_1 + 1/R_2}$ where σ_s is the SD of a single run and R_1 and R_2 are the number of the runs in the respective samples compared. This gives the SD of the difference for run score averages.

SERIES: Several runs or experimental sessions that are grouped in accordance with a stated principle.

SESSION: A unit of an ESP or PK experiment comprising all the trials of one test occasion.

SET: A subdivision of the record page serving as a scoring unit for a consecutive group of trials, usually for the same target.

SEVENS TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling seven.

•SIGNIFICANCE: A numerical result is significant when it equals or surpasses some criterion of degree of chance improbability. Common criteria are: a probability value of .01 or less, or a deviation in the expected direction such that the critical ratio is 2.33 or greater.

SINGLES TEST: A PK technique in which the aim of the subject is to try to influence dice to fall with a specified face up.

STIMULUS OBJECT: The ESP card or drawing or other object, some identifying characteristic of which is to be apprehended by the percipient.

TARGET: In ESP tests, the stimulus object; or in telepathy, the mental state of the agent. In PK tests, the faces of the die (or combination of faces) which the subject attempts to bring up in the act of throwing.

TARGET CARD: The card which the percipient is attempting to perceive (i.e., to identify or otherwise indicate a knowledge of).

SUBJECT: The person who is experimented upon. In ESP tests, most commonly the percipient (though also the agent in GESP and telepathy). In PK tests, any individual whose task it is to influence the objects thrown.

TARGET DECK: The deck of cards the order of which the subject is attempting to identify.

TARGET FACE: The face on the die which the subject tries to turn up as a consequence of direct mental action.

TELEPATHY: Extrasensory perception of the mental activities of another person. It does not include the clairvoyant perception of objective events.

TRIAL: In ESP tests, a single attempt to identify a stimulus object. In PK tests, a single throw of the dice or other objects thrown.

Dr. Charles E. Stuart